

Classification and Prediction on Rural Property Mortgage Data with Three Data Mining Methods

Kaixi Zhang, Yingpeng Hu, Yanghui Wu*

School of Science, Northwest A&F University, Yangling, China

Email: *sxxawyh08@nwsuaf.edu.cn

How to cite this paper: Zhang, K.X., Hu, Y.P. and Wu, Y.H. (2018) Classification and Prediction on Rural Property Mortgage Data with Three Data Mining Methods. *Journal of Software Engineering and Applications*, 11, 348-361.

<https://doi.org/10.4236/jsea.2018.117022>

Received: May 15, 2018

Accepted: July 21, 2018

Published: July 24, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Farmers Property Mortgage Policy is a strategic financial policy in western China, a relatively underdeveloped region. Many contradictions and conflicts exist in the process between the strong demand for the loans by farmers and the strict risk control by the financial institutions. The rural finance corporations should use scientific analysis and investigation of the potential households for overall evaluation of the customers. These include historical credit rating, present family situation, and other related information. Three different data mining methods were applied in this paper to the specifically-collected household data. The objective was to study which factor could be the most important in determining loan demand for households, and in the meanwhile, to classify and predict the possibility of loan demand for the potential customers. The results obtained from the three methods indicated the similar outputs, income level, land area, the way of loan, and the understanding of policy were four main factors which decided the probability of one specific farmer applying for a credit loan. The results also embodied the difference within the three methods for classifying and predicting the loan anticipation for the testing households. The artificial neural network model had the highest accuracy of 91.4 which is better than the other two methods.

Keywords

Rural Property Mortgage, Bayesian Network, Artificial Neural Network, Logistic Regression, Classification and Prediction

1. Introduction

As a developing country, unbalanced development exists in China. The development of western China is far behind that in the eastern part, especially in the

rural areas of the western China. The central government encourages the rural financial cooperatives to provide loans to farmers in order to expand the scope of production. However, the potential credit risks constrain the credit operation between commercial banks and households. All the financial institutions need to seriously examine the basic situation and business of each household to decide the probability of lending. Rural and agricultural development is facing a bottleneck caused by insufficient funds and lack of financial credit. The national banks and rural financial cooperatives tighten lending strictly due to concern that the risk can not be controlled. The waiting time for a loan is invariably very long.

From experience and common sense, we know that some factors exist which restrict the enthusiasm of households to apply for credit loan from financial institutions. For example, the various educational backgrounds of the household's head lead to the different understandings of the loan policies. The household land area determines a family's income level and consumption level. The latter in turn causes different demands for loans. On the other hand, the financial institutions need to consider the farmer's credit risk through investigating the complex dependencies from loan history, educational level, proportion of income and consumption, family size and other factors of the specific family.

How to investigate the causal relationships among all these factors and how to take advantage of them to predict the probability of one specific farmer will be of practical value. Predicting the possibility of loan demand is one of the most interesting and challenging tasks in which to develop data mining applications. With the increased use of computing methods and data mining techniques, large volumes of financial data are being collected and are being made available to the specific research community. Prediction models are being developed with these historical data based on knowledge discovery methods such as statistics or other optimization techniques. All of these models identify and exploit relationships among large numbers of variables regarding households and financial institutions, and are able to predict the outcome of loan demand using the historical cases stored within a database.

Previous researches utilized statistical models to study the correlative analyses of all these factors on the loan demand [1] [2] [3]. In rural financial domains, where data and statistics driven research is successfully applied. The results could produce some useful suggestions for financial institutions to evaluate the specific household. For instance, in [1], a multivariate regression model, a Logistic regression model is applied to study the relationship between the independent variable, loan demand and other variables and influencing factors. However, we know that all statistical models are based on some assumption. For example, all variables considered must be subject to the normal distribution conditions in addition to the assumption that the variables are mutually independent. What is lacking of this assumption is that it means the statistical results will be accepted with a certain confidence value.

Causal relationships among variables can provide intuitive observation for a particular household and can provide support for financial institutions to make

a scientific assessment. There are many popular data mining methods that can be used to study these specific problems [4] [5]. In this paper, we utilise two machine learning methods (Bayesian network and Artificial neural network) and one statistical technique (Logistic regression) to build models for investigating the interested variables, and to develop prediction models for household loan demands. We have designed a questionnaire and gathered a general case of data for consideration, where the observations are discrete. The objective of this paper is to study the causal relationships among variables with three data mining methods in rural financial data, to build and evaluate the classification and prediction models under the three methods. We expect the empirical analysis and the theoretical results can provide a valuable reference to the relevant financial institutions when they are assessing credit loan. The innovation of this paper lies in the application of two typical data mining methods (Bayesian Network and Artificial Neural Network) to predict and analyze the data of farmers to overcome the insufficiency of traditional statistical model (Logistical Regression) Practical significance £ The work of this paper is the chief application of data mining method in the prediction of economic data in western China and the research results of this paper have certain reference significance for the analysis of rural financial mortgage loan policies in western China.

The paper is organized as follows: we introduce the data and their properties in Section 2. In Section 3, we present the three methods respectively. The comparative analysis of classifications and predictions is described in Section 4. The conclusion is summarized in Section 5.

2. Datasets

The data used in this paper was collected during June 2011 and July 2012 by the researchers from College of Economics and Management, Northwest A&F University, China. The whole data collection process was supported through funding from the Chinese government. The project is “Changjiang Scholars and Innovative Research Team in University, Jan 2012-Dec 2014, No.IRT1176”. All these data were taken from the western region of China, including Shaanxi province and Ningxia province. In order to ensure that the data is scientific and reasonable, we randomly surveyed a total of 4000 households from the above regions using a questionnaire. The data collected consists of three main parts. The first part is composed of basic information relating to the specific investigated farmer, including age, educational level, family size, land management, household income and expenditure structure, etc. The second part includes loan status of the farmers, loan history in the past 5 years and credit rating. The third part is made up of the understanding, demand and satisfaction about the property rights mortgage. We selected a total number of 11 factors in this research, each of these factors has 2 to 5 attributes to describe the different levels of the specific household. For example, the variable “Income (CNY)” has 5 levels (0 – 5000, 50001 – 10000, 10001 – 20000, 20001 – 50000, 50000 +) according to the

true income of the household. All the data considered in this paper are listed in **Table 1**.

The meaning of each variable in 1 is described as following: Income represents the income level of the specific household, Wayofloan is for the way a specific household ever used, Expenditure is for the spending level of a household, FamilySize is the population living in a household, LoanDem and describes if the household need a loan or not, Policy means the level to which a household understands the loan policy, Land Area is the land size a household owns, Age is the true age of the householder, Edu is the educational background of the householder, and Conven is how easy it is for a household to apply for a loan.

Table 1. Data description of rural household in western china.

Factor	Level	Attribute	Factor	Level	Attribute
Income (CNY)	1	0 - 5000	Policy	1	Never heard
	2	5001 - 10,000		2	Heard a little
	3	10,001 - 20,000		3	General understanding
	4	20,001 - 50,000		4	Basic understanding
	5	50,000+		5	Know very well
WayofLoan	1	Formal institution	Land Area (667 m ²)	1	0 - 3
	2	Private lending		2	4 - 10
	3	Never lend		3	10+
Expend (CNY)	1	0 - 5000	Age	1	Under 29
	2	5001 - 10,000		2	30 - 39
	3	10,001 - 20,000		3	40 - 49
	4	20,001 - 50,000		4	50 - 59
	5	50,000+		5	60+
FamilySize	1	1 - 3	Edu	1	Unschoolled
	2	4 - 5		2	Primary school education
	3	5+		3	middle school education
LoanDemand	1	No demand	Conven	4	high school education
	2	Demand		5	college educated
Gender	1	Male	Conven	1	Inconvenient
	2	Female		2	Convenient

3. Methods and Prediction Models

In keeping with recently published literature as well as our previous studies, we will take three different types of classification models in this paper. They are the Bayesian Network Model (BN), Artificial Neural Networks (ANN), and Logistic Regression (LR). A simple introduction of these models is as follows:

3.1. Bayesian Network Model

Bayesian Networks (BNs) are probabilistic graphical models which represent the dependencies among a set of random variables in a chosen domain [6] [7]. A BN structure consists of two main components: a visible Directed Acyclic Graph (DAG) and a set of parameters. The DAG is defined as $G = (V, \varepsilon)$, V is the node set and ε is the edge set. Each node X_i represents a random variable relevant to the problem domain. Each edge connects attributes that are directly dependent on a pair of nodes. If there is an arrow from node X_j to node X_i , it means that X_j is a parent of X_i , or equivalently, X_i is the child of node X_j . The set of parents of a node X_j is denoted as $p(X_j | P_a(X_j))$. Thus, the conditional distributions between node X_j and its parents set $P_a(X_j)$ is denoted as $P_a(X_j)$. The dependencies among these random variables $X = \{X_1, X_2, \dots, X_n\}$ are represented as the joint probability distribution (JPD)

$$p(X) = \prod_{i=1}^n p(X_i | P_a(X_i)) \quad (1)$$

Because a Bayesian network is a complete model for the variables and their relationships, it can be used for Bayesian inference. For example, the network can be used to predict the probability of a state to any interested variable when other variables are observed. This process means to compute the posterior distribution of the variable given evidence along with the prior probability and the specific BN structure. When a BN structure is built, one can use this model as an expert system to get this posterior probability through applying the Bayes' theorem to the complex problem [8] [9].

3.2. Artificial Neural Network

Artificial Neural networks (ANNs) are commonly known as biologically inspired analytical techniques, capable of predicting new observations from other observations after executing from existing data. ANNs are basically a data-driven black-box model to explore the relationships between input and output variables from historical data. They are virtual input-output device that accept any number of numeric inputs and produce any number of numeric outputs. ANNs have the ability to solve highly non-linear complex problems [10] [11] [12]. Multi-layer perceptron (MLP) with back-propagation is a popularly used and well-studied ANN model. It is known as a powerful function approximator for prediction and classification problems. In this paper, a three-layer feed forward ANN will be applied. The model has 10 input neurons in the input layer, many hidden neurons and two neurons in the output layer.

3.3. Logistical Regression

Logistical regression (LR) is a generalization of linear regression [13] [14]. It is developed for analysing data with categorical dependent variables, especially when there are only two categories of the dependent variable. Because the dependent variable is discrete, it cannot be modeled directly by linear regression

and cannot predict a numerical value using LR. With binomial probability theory, in a two-class problem, LR builds a model to predict the odds of each event's occurrence rather than predicting the point estimate of the event itself. For example, odds greater than 0.5 means that the case is assigned to one group rather than the other. LR generates a best fitting equation using maximum likelihood method. The objective of this equation is to maximize the probability of classifying the observed data into the appropriate category given the regression coefficients. LR can be used for predicting the group membership depending on the results of an odds ratio. On the other hand, it can provide knowledge of the relationships and strengths among the variables.

4. Classification, Prediction Results and Discussion

In this section, we first carry out the relationships analysis within factors with BN, ANN and LR respectively. The results of the comparison of these outputs provide factor classification in different perspectives. We then study the accuracy of each model with testing data. The properties of accuracy about these models embody the authenticity and reliability when they are utilised in real problems. In the first part, we randomly select half of the total data size (2000 cases) as training data set for building the classification model. The rest of the data (2000 cases) are adopted as testing data set to test each of the models and assess the accuracy of each model.

4.1. Relationships Analysis and Classification

4.1.1. BN Classification Results

We took an novel algorithm, ChainACO, in this paper for BN topological graph learning. ChainACO is an algorithm which is developed by Wu, *et al.* [15] [16]. It has been tested as an efficient and cheap technique for BN structure learning, especially for large groups of data. In this problem, we run ChainACO and achieved the best structure accompanying the best fitness score within 10 repeated runs. The structure obtained is described in **Figure 1**.

Figure 1 reveals a topological graph about all the controlled variables in this study. In this structure, variable LoanDemand is the key factor which is more greatly concerned about by the financial institutions. The BN structure indicates that there are four factors directly influencing the demand of loan for a specific household. They are Income, LandArea, Policy and the WayofLoan. The BN structure also exposes the relationships between these four variables, for instance, Income is the parent of Expend, the Expend effects the variable Policy and WayofLoan. However, the WayofLoan is the parent of variable Policy, *et cetera*. In addition, this topological graph describes other useful results, such as the two factors, Conven and Edu are dependent to the other 9 factors. This means both the Conven and educational background of one specific household can be ignored when we consider the loan demand. The BN structure achieved from financial data indicates the underlying relationship with all factors in this

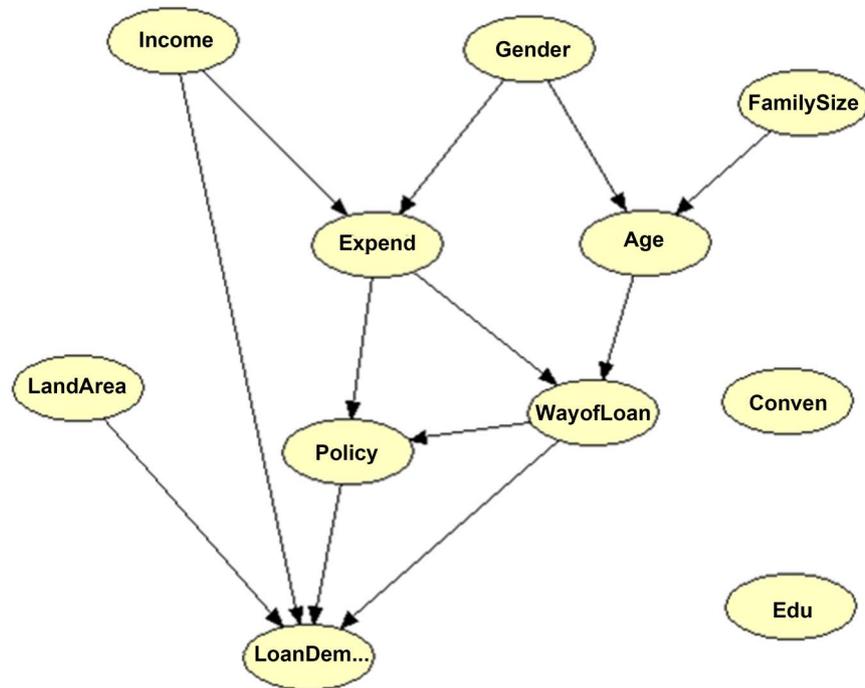


Figure 1. Bayesian Network structure achieved with ChainACO algorithm.

problem. The visualization of the results provides primary and intuitive reference and suggestions for the Bank organizer when making loan policy.

On the other side, the above BN model can provide quantitative relationships to any one interested variable and the relevant variables. For example, **Table 2** is the conditional probability distribution among the variables LoanDemand, Policy, Income, LandArea and WayofLoan. This table shows to what degree the basic information of each farmer restricts the possibility of applying for a loan from the financial institutions, and which combination of this basic information about the farmer makes him the most possible customer for applying for the loan. For instance, the farmer who knows the loan policy very well, has an income level in level 1, with land area in level 2, and takes the loan in way 1 will have the highest possibility to apply for the loan in the future (the probability in this case is 0.6190). In the BN model, we can gather a conditional probability distribution about any one factor with its corresponding factors. **Table 3** is an other example about the factor Wayofloan related to the relevant variables, Expend and Age. **Figure 1** shows that the variables Expend and Age are parent factors to Wayofloan, so the different combinations of specific farmers have different attitude to the credit loan.

BN provides us with a visual topological graph that indicates the underlying relationships among all interested factors. The quantitative conditional probability distribution reveals the inherent probabilistic relationships of these factors.

4.1.2. Logistical Regression Results

We use a popular statistics tool, SPSS 21 to carry out the logistic regression analysis [17]. In this process, the factor LoanDemand is regarded as the

Table 2. BN Probabilities.

LandArea		2	2	2
Policy		5	2	2
WayofLoan		1	1	1
Income		1	4	5
LoanDemand	1	0.6190	0.2065	0.0476
	2	0.2810	0.6932	0.9524

Table 3. BN Probabilities.

Expend		2	2	3
Age		1	2	2
WayofLoan	1	0.6000	0.8269	0.4929
	2	0.0444	0.0192	0.0286
	3	0.3555	0.1538	0.0286

dependent variable, and all the other variables are independent factors. We try to study the relationships between this dependent variable and all other variables, trying to investigate the significance of this effect. The results about variables in the equation and the variables not in the equation along with the corresponding significance level are listed in **Table 4**.

In **Table 4**, Wald and Sig. indicates the Wald chi-square test that is used to test the constant and an independent variable would be significant in the model. B means the coefficient for the constant or the independent variables if they are statistically significant to the model. These coefficient are the values in the logistic regression equation for predicting the dependent variable from the independent variable. Looking at the Sig. values in **Table 4**, we can see that variables: Income, WayofLoan, Expend, Policy, and LandArea are statistically significant in the model (p-value is smaller than the critical p-value of 0.05), the rest of the variables are not statistically significant.

According to the above conclusions, the logistic regression equation is Equation (2)

$$P(\text{Loan Demand}) = \frac{1}{1 + \exp(-(-2.602 + 1.310 * X_1 - 3.059 * X_2 + 0.771 * X_3 + 3.063 * X_4 + 4.074 * X_5))} \quad (2)$$

In Equation (2), X_1 , X_2 , X_3 , X_4 , and X_5 represents Income, Wayof-Loan, Expend, Policy and Land Area respectively. The regression equation tells us about the relationship between the independent variables and the dependent variable. These estimates show the amount of increase (or decrease, if the sign of the coefficient is negative) in the predicted log odds of Loan Demand = 1 that would be predicted by a 1 unit increase (or decrease) in the predictor, holding all

Table 4. Experimental results with Logistical Regression on financial data.

Variables	Variables in equation			Variables not in equation		
	B	Wald	Sig.	Variables	Score	Sig.
Income	1.310	26.631	0.000	Family Size	0.057	0.811
WayofLoan	-3.059	240.350	0.000	Gender	1.199	0.274
Expend	0.771	0.246	0.002	Age	0.090	0.764
Policy	3.063	150.385	0.000	Edu	0.031	0.859
LandArea	4.074	0.302	0.000	Conven	3.662	0.056
Constant	-2.602	0.280	0.000			

other predictors constant. For instance, for every one-unit increase in Income score, we expect a 1.310 increase in the log-odds of Loan Demand, holding all other independent variables constant. The LR equation produced the correlation coefficient $R^2 = 0.806$, The value is significant at the 0.01 level (2-tailed).

4.1.3. Artificial Neural Network Classification on Rural Data

Artificial Neural Network was performed using SPSS 21. In order to build the structure of ANN, the training data were randomly assigned to training (1398 cases; 69.5%) and testing (602 cases; 30.5%) datasets. The input layers consisted of ten input nodes, and the output layer has one node with two states (Loan Demand = 1 and Loan Demand = 2). After the debugging and testing five times, in this research, the hidden layer consisted of 10 hidden nodes.

The first main result produced in this model is the importance of each input factor to the dependent variable, which is shown in **Figure 2**. This table depicts the three main variables: Land Area, WayofLoan, and Policy respectively. The corresponding respective relevance of the three variables are significant importance to other seven variables. The fourth important variable is Income. It has a lower impact factor, only about 30% that of other factors, and the normalized importance is less than 20%, these suggest that they are negligible when considering the impacts to the loan demanding of one specific farmer.

The second output which we concerned is the correlation coefficient between actual data and estimated values. The ANN model produced the correlation coefficient $R^2 = 0.806$, The value is significant at the 0.01 level (2-tailed). The result demonstrates that the classification is reliable for the specific dataset.

Comparison of the above three methods for dataset analysis and classification shows the common conclusion that they all can perform specific classification and draw the main factors which affect the dependent variable, Loan Demand. In spite of the different performance in each method, they all demonstrated that the Income, Landarea, Policy and WayofLoan are the most important four factors relating to the key factor, Loan Demand. Furthermore, BNN provides the potential direct and indirect relationships among these factors and the other factors. LR reported the statistical correlation (including positive and negative

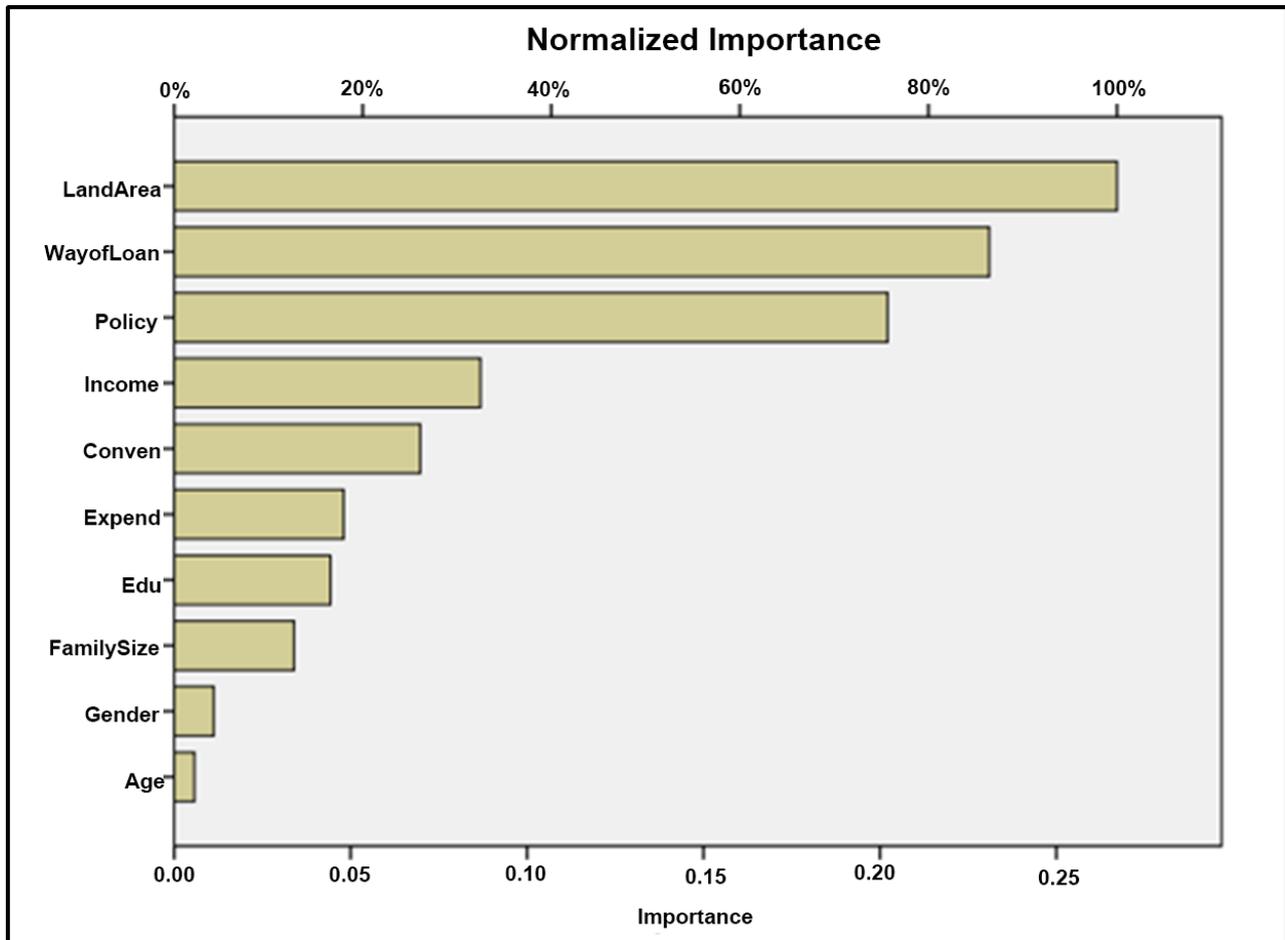


Figure 2. Importance and Normalized importance of each factor in ANN model.

correlations) of each factor to the Loan Demand. ANN presented the importance in quantitative terms with each factor to the dependent factor.

The investigation of multi-factors analysis is a popular problem in rural finance. Comprehensively comparing these results can provide us with inspiration for understanding the substantial problem in rural finance. For instance, the data in this paper is collected from the less developed regions in western China. In these regions, the farmer's main income comes from land they have owned, more land that they own, and higher incomes from agriculture. So the investment of household's land had significantly positive effect on the credit loan. From the result in LR, we can see that the Land Area has the highest positive coefficient (4.074) to the factor, Loan Demand. In ANN model, Land Area is the most critical influence to Loan Demand. Educational background should be an important factor when applying for a loan, however, all the classification results show it is a weaker factor in this problem. For instance, in BN, the Edu is independent to all other factors, in LR, it is not included in the equation (the Sig. value is 0.859), and in ANN, the Normalized importance is less than 20%. The explanation for this performance is that in rural areas with a lower level education development, farmers who intend to apply for the credit loan depend on the

actual need but not on the education degree they have. The valuable suggestion is with synthesizing the outputs of the above methods, the useful results can be concluded for analysing and investigating the collected data in rural area as discussed.

4.2. Measures for Classification and Prediction Results

Sensitivity, specificity, and accuracy, are widely used statistics to describe a prediction and classification model. They are used to quantify how good and reliable a classification is. In this rural financial problem, sensitivity evaluates how good the classifying is at detecting a positive result. Specificity estimates how likely it is that a farmer who does not need a loan can be correctly ruled out. Accuracy measures how correctly a classification identifies and excludes a given condition [18] [19] [20]. Sensitivity, specificity, and accuracy are described as follows

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

In these equations, TP , TN , FN and FP mean true positive, true negative, false negative and false positive respectively. For example, if a household really needs a credit loan from financial institution, and the given classification also indicates that the farmer needs a loan, the result of the classification is considered true positive (TP). Similarly, if a household does not need credit, and the classification result shows the same one, the test result is true negative (TN). Both true positive and true negative suggest a consistent result between the classification and the truth. If the classification model confirms a household does not need a loan, but the household actually does want the credit loan, the test result is false positive (FP). Similarly, if the result of the classification suggests the farmer needs a loan, but he actually does not need it, the test result is false negative (FN). Both false positive and false negative indicate that the classification results are opposite to the actual condition.

We apply the models construed in the previous section to the testing data in our problem. We have understood the basic information about all these data. For example, we know the situation of loan demand for each household. Through comparing the prediction results about Loan Demand to the actual Loan Demand, we got TP , TN , FN and FP to calculate the sensitivity specificity and accuracy. **Table 5** describes the comparison results of accuracy, sensitivity and specificity within these three methods.

Table 5 shows the complete set of results in a tabular format. For each model, the detailed prediction results of the validation datasets are presented in form of confusion matrixes. The problem in this paper is a two-class prediction problem,

Table 5. Comparison of Accuracy within three methods.

	Confusion Matrix		Accuracy	Sensitivity	Specificity
BN	437	146	0.880	0.750	0.934
	94	1322			
ANN	448	135	0.914	0.769	0.973
	410	1375			
LR	453	130	0.907	0.777	0.960
	56	1360			

the upper left cell denotes the number of samples classified as true while they were true (TP), and lower right cell denotes the number of samples classified as false while they were actually false (TN). The lower left cell and upper right cell denote the number of samples misclassified. So the lower left cell is FP and the other one is FN. Once the confusion matrixes were build, the accuracy, sensitivity and specificity of each fold were calculated using the respective formulas presented in the previous section.

In evaluating the performance of the above three methods, we found that the BN achieved a classification accuracy of 0.88 with a sensitivity of 0.75 and a specificity of 0.934. The LR model achieved a classification accuracy of 0.907 with a sensitivity of 0.777 and a specificity of 0.960. However, the ANN performed the best among the three models evaluated. It achieved a classification accuracy of 0.914 with a sensitivity of 0.769 and a specificity of 0.973. In the three different classification models, the numerical values of specificity are all bigger than 0.90, which means that when we assess a farmer who does not need a loan, there is a very high chance (for instance, 0.97 with ANN result). The numerical values of sensitivity indicate the probability that an assessment identifies farmers as needing credit loans who do in fact need it. However, these values are only close to 0.80 suggesting that the financial institutions should give more deep investigation to these potential customers for financial security.

5. Conclusions

In this paper, we report a research effort where we developed three prediction models for farmers loan demand. Two of the models are from machine learning (BN, ANN) and one from statistics (LR). All of these methods are introduced to study the large dataset (4000 cases with 11 factors) which we investigated from western China. Each of these methods reveals the potential relationships among factors from different perspectives. BN embodies these relationships through visual graph and quantitative table, ANN mines the importance of each factor to the dependent factor. However, the LR exploits the statistical correlation about these factors. In spite of difference found in the three methods, the results should provide valuable suggestions for understanding the loan policy of the corresponding national financial organization.

The Accuracy measure for each prediction model is calculated. The ANN results indicate that it performed the best with a classification accuracy of 91.4, which is better than the other two methods. These results suggest that each model can be developed to accurately predict the outcome of a farmer with loan demanding. The ANN model can be valuable tool in the rural finance industry in China. The model can be used to assist in making financial policy, developing the rural economy.

The complexity, validity, as well as the accuracy of data directly decide the classification efficiency and predication accuracy. BN, ANN and LR require large amount of training data. Also, the accuracy in this paper still needs to be further improved. A larger data set and the improved Bayesian network or neural network will be used to improve the accuracy in the future. Our ongoing research efforts are geared toward investigating large data set from western China and studying properties of these methods.

Acknowledgements

This paper is partially supporting by programs for the Fundamental Research Funds for the Central Universities (2452015223), the Scientific Research Foundation for doctorate of Shaanxi Province of China (Z111021504), and the Scientific Research Foundation for doctorate of Northwest A & F University (Z111021306).

References

- [1] Zhang, Y.-Y., Wang, F. and Luo, J.-C. (2013) An Empirical Study on the Influence Factors of Households' Normal Credit Default. *Economic Survey*, **1**, 33-37. (In Chinese)
- [2] Cheng, L.F. (2005) Analysis on Chinese Urban Households' Economic Structure. Chinese Economy Press. (In Chinese)
- [3] Hong, H.L., Bai, Y. and Wu, Y.L. (2011) Credit Rationing in Local Financial Development. *Economic Management*, **2011**, 147-154. (In Chinese)
- [4] Battiato, S., Farinella, G.M. and Giuffrida, G. (2010) Exploiting Visual and Text Features for Direct Marketing Learning in Time and Space Constrained Domains. *Pattern Analysis and Applications*, **13**, 143-157.
<https://doi.org/10.1007/s10044-009-0145-2>
- [5] Borgelt, C., Steinbrecher, M. and Kruse, R.R. (2009) Graphical Models—Representations for Learning, Reasoning and Data Mining. 2nd Edition, John Wiley & Sons, Hoboken.
- [6] Larranaga, P. and Moral, S. (2011) Probabilistic Graphical Models in Artificial Intelligence. *Applied Soft Computing*, **11**, 1511-1528.
<https://doi.org/10.1016/j.asoc.2008.01.003>
- [7] Li, Z. (2013) Systematically Discovering Dependence Structure of Global Stock Markets Using Dynamic Bayesian Network. *Journal of Computational Information Systems*, **9**, 7215-7226.
- [8] Fenton, N. and Neil, M. (2012) Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press, Boca Raton.

- [9] Yan, Y. and Suo, B. (2013) Risks Analysis of Logistics Financial Business Based on Evidential Bayesian Network. *Mathematical Problems in Engineering*, **2013**, 1-8. <https://doi.org/10.1155/2013/416282>
- [10] Ayat, S., Farahani, H.A., *et al.* (2013) A Comparison of Artificial Neural Networks Learning Algorithms in Predicting Tendency for Suicide. *Neural Computing and Applications*, **35**, 1381-1386. <https://doi.org/10.1007/s00521-012-1086-z>
- [11] Ogwueleka, F., Misra, S., *et al.* (2015) Neural Network and Classification Approach in Identifying Customer Behavior in the Banking Sector: A Case Study of an International Bank. *Human Factors and Ergonomics in Manufacturing & Service Industries*, **25**, 28-42.
- [12] Ankinakatte, S., Norberg, E., Lovendahl, P., Edwards, D. and Hojsgaard, S. (2013) Predicting Mastitis in Dairy Cows Using Neural Networks and Generalized Additive Models: A Comparison. *Computers and Electronics in Agriculture*, **5**, 1-6.
- [13] Hosmer, D.W. and Lemeshow, S. (2000) Applied Logistic Regression. 2nd Edition, John Wiley & Sons, New York.
- [14] Miyamoto, M. (2014) Credit Risk Assessment for a Small Bank by Using a Multinomial Logistic Regression Model. *International Journal of Finance and Accounting*, **3**, 327-334.
- [15] Wu, Y., McCall, J.A.W. and Corne, D.W. (2010) Comparative Analysis of Search and Score Metaheuristics for Bayesian Network Structure Learning Using Node Juxtaposition Distributions. Springer, New York, 424-433.
- [16] Wu, Y., McCall, J.A.W. and Corne, D.W. (2010) Two Novel Ant Colony Optimization Approaches for Bayesian Network Structure Learning. *IEEE Congress on Evolutionary Computation*, Barcelona, 18-23 July 2010, 1-7.
- [17] George, D. and Mallery, P. (2013) IBM SPSS Statistics 21 Step by Step: A Simple Guide and Reference. 13th Edition, Pearson, London.
- [18] Perzyk, M., Kochanski, A., Kozłowski, J., *et al.* (2014) Comparison of Data Mining Tools for Significance Analysis of Process Parameters in Applications to Process Fault Diagnosis. *Information Sciences*, **5**, 380-392.
- [19] Duda, R.O., Hart, P.E. and Stork, D.G. (2001) Pattern Classification. 2nd Edition, John Wiley & Sons, New York.
- [20] Egmont-Petersen, M., de Ridder, D. and Handels, H. (2002) Image Processing with Neural Networks? A Review. *Pattern Recognition*, **5**, 2279-2301. [https://doi.org/10.1016/S0031-3203\(01\)00178-9](https://doi.org/10.1016/S0031-3203(01)00178-9)