MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 1401-1412

# The Logit Regression And Neural Network Analysis Done By Reliance Group

**Surya.S[1]*, Dr. Asha Sundaram[2], Dr. Thangamayan [3]**
[1]Research scholar -saveetha school of law,SIMATS,Chennai
[2]Principal& Professor -Saveetha School of law,SIMATS,Chennai
[3]Associate Professor&HOD -Saveetha School of law,SIMATS,Chennai

**File Name: loan sanction**

US Small Business Administration is helping small businesses by lending and guaranteeing a small portion of the loan. You will then want to classify this loan as "higher risk—more likely to default" or "lower risk—more likely to not default" when making your decision. The attributes are given below:

**Attribute Information:**

| Variable Name | Data Type | Description of variable |
|---|---|---|
| Noemi | Number | Number of Business Employees |
| New | Text | 1 = Existing Business, 2 = New Business |
| Create jobs | Number | Number of jobs created |
| UrbanRural | Text | Location: 1= Urban, 2= Rural, 0 = Undefined |
| DisbursementGross | Currency | Amount Disbursed |
| Portion | Number | The proportion of gross amount guaranteed by the Government when sponsoring the business |
| RealEstate | | =1 if the loan is backed by real estate, =0 otherwise |
| Recession | | =1 if the loan is active during the Great Recession, =0 otherwise |

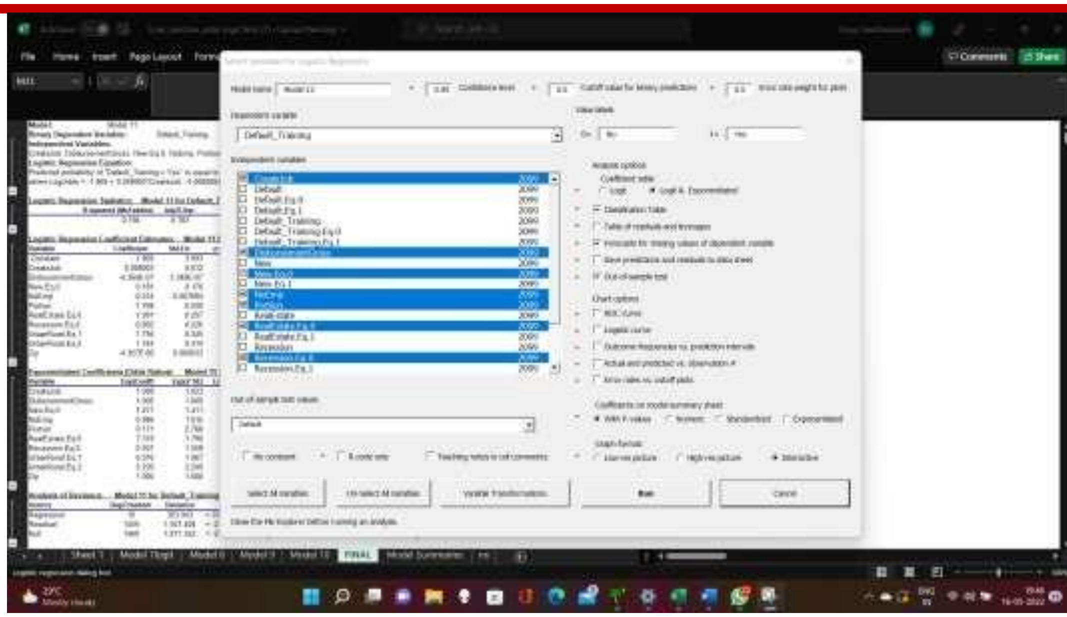
1. Build a logistic regression model to classify "default" (Dataset: loan_sanction_data- logit.xlsx")

a. The logit regression model includes the steps such as

i. Step 1- load the data in regression excel software
ii. Load the data
iii. Try to use the icons select data and add names
iv. The click the logit regression icon
v. The regression box pops up
vi. Select the dependent variable as default training \
vii. Select the transformation variable and add a dummy variable to real estate, employment, default, new job
viii. After creating the dummy variable transformation with respect to K-1 condition select the variable
ix. Then in out of sample try to select the default variable
x. The run the data
xi. \the end of regression model

A) Interpret the value obtained. and finalize the data.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. xx Issue x, Jan-June 20yy, Pages: 357-394



2. How did you improve the predictive power of the model?

- The predictive model mainly helps in identifying who is going to churn. It is mainly interested in predicting accuracy
- Model performs concerning the accuracy
- To remove the insignificant value in the model
- P-value mainly explains about the sample matches the population, the goal of the predictive model is to build the training data concerning test if it works with the test data

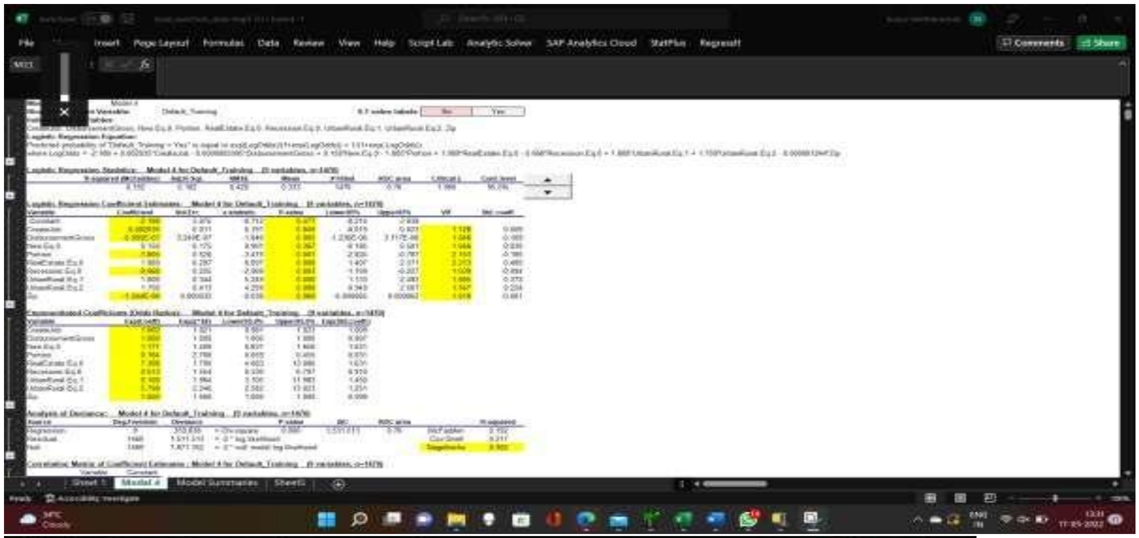
- Three criteria with which the predictive model works is

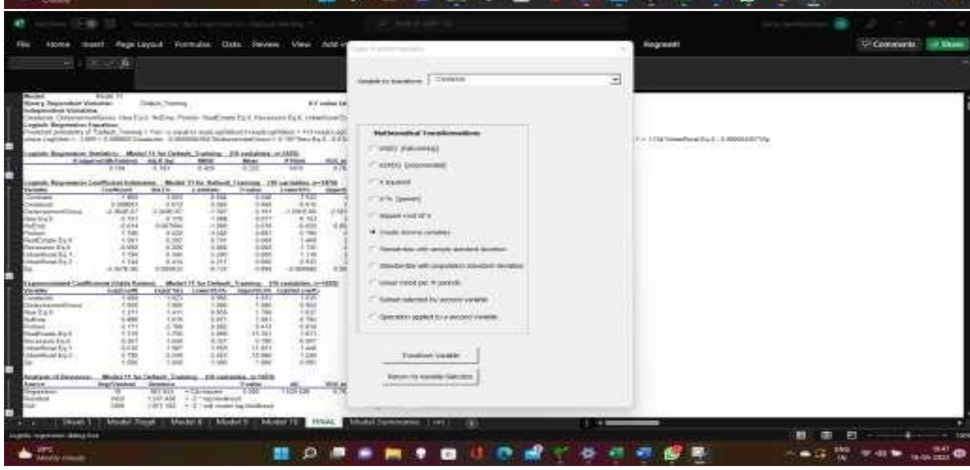
  a. Specificity
  b. Accuracy
  c. Sensitivity
  d. Considering the cut-off value

- For each record model calculates the probability if greater than the cut-off value then the positive class or negative class can be classified
- If we check with training data and perform poorly with test it is called overfitting, memorized training data of out of sample.
- If the cut-off value is been adjusted in the confusion matrix then we can obtain a good fitting model to further determine the predictive nature of data using regression
- If the model is overfitting then use the stratified sampling, and adjust the cut off rate else try to gather data furthermore to make the model consistent and fitting.

In this loan section data set, the data in the confusion matrix obtained is fitting and we decided to adopt this model as it has a more true negative rate with a 0.50 cut-off that is specificity

**MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal**
**ISSN: 1053-7899**
**Vol. xx Issue x, Jan-June 20yy, Pages: 357-394**





- To analyze the factors that are related to the output in the regression model are
  - VIF should not be greater than 5
  - The coefficient should be lesser than 0.05
  - P-value is pseudo R square value should be lesser than 0.198
  - The nagelkerke value should vary above 20%
  - The exponential coefficient are classified based on
    - Greater than 1 the loan is accepted
    - Equal to 1 the equal chances for sanctioning the loan and rejecting the loan
    - Lesser than 1 rejecting the loan

While analysing the output obtained from the logit regression model we could infer that those variables that are influencing the default training data are

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. xx Issue x, Jan-June 20yy, Pages: 357-394

Logistic Regression Coefficient Estimates: Model 4 for Default_Training (9 variables, n=1470)

| Variable | Coefficient | Std.Err. | z-statistic | P-value | Lower95% | Upper95% | VIF | Std. coeff. |
|---|---|---|---|---|---|---|---|---|
| Constant | -2.188 | 3.075 | -0.712 | 0.477 | -8.214 | 3.838 | | |
| CreateJob | 0.002035 | 0.011 | 0.191 | 0.849 | -0.019 | 0.023 | 1.128 | 0.009 |
| DisbursementGross | -5.995E-07 | 3.249E-07 | -1.846 | 0.065 | -1.236E-06 | 3.717E-08 | 1.648 | -0.109 |
| New.Eq.0 | 0.158 | 0.175 | 0.901 | 0.367 | -0.185 | 0.501 | 1.046 | 0.030 |
| Portion | -1.805 | 0.520 | -3.475 | 0.001 | -2.824 | -0.787 | 2.153 | -0.185 |
| RealEstate.Eq.0 | 1.989 | 0.297 | 6.697 | 0.000 | 1.407 | 2.571 | 2.213 | 0.489 |
| Recession.Eq.0 | -0.668 | 0.225 | -2.969 | 0.003 | -1.109 | -0.227 | 1.039 | -0.094 |
| UrbanRural.Eq.1 | 1.808 | 0.344 | 5.249 | 0.000 | 1.133 | 2.483 | 1.665 | 0.372 |
| UrbanRural.Eq.2 | 1.758 | 0.413 | 4.259 | 0.000 | 0.949 | 2.567 | 1.547 | 0.224 |
| Zip | -1.244E-06 | 0.000032 | -0.038 | 0.969 | -0.000065 | 0.000062 | 1.018 | -0.001 |

Exponentiated Coefficients (Odds Ratios): Model 4 for Default_Training (9 variables, n=1470)

| Variable | Exp(Coeff) | Exp(z*SE) | Lower95.0% | Upper95.0% | Exp(Std.coeff.) |
|---|---|---|---|---|---|
| CreateJob | 1.002 | 1.021 | 0.981 | 1.023 | 1.009 |
| DisbursementGross | 1.000 | 1.000 | 1.000 | 1.000 | 0.897 |
| New.Eq.0 | 1.171 | 1.409 | 0.831 | 1.650 | 1.031 |
| Portion | 0.164 | 2.768 | 0.059 | 0.455 | 0.831 |
| RealEstate.Eq.0 | 7.308 | 1.790 | 4.083 | 13.080 | 1.631 |
| Recession.Eq.0 | 0.513 | 1.554 | 0.330 | 0.797 | 0.910 |
| UrbanRural.Eq.1 | 6.100 | 1.964 | 3.105 | 11.983 | 1.450 |
| UrbanRural.Eq.2 | 5.799 | 2.246 | 2.582 | 13.023 | 1.251 |
| Zip | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 |

Analysis of Deviance: Model 4 for Default_Training (9 variables, n=1470)

| Source | Deg.Freedom | Deviance | | P-value | AIC | ROC area | | R-squared |
|---|---|---|---|---|---|---|---|---|
| Regression | 9 | 359.838 | = Chi-square | 0.000 | 1,531.513 | 0.76 | McFadden | 0.192 |
| Residual | 1460 | 1,511.513 | = -2 * log likelihood | | | | Cox-Snell | 0.217 |
| Null | 1469 | 1,871.352 | = -2 * null model log likelihood | | | | Nagelkerke | 0.302 |

The portion- the role of government sanction the loan plays a main variable concerning the +1 increase

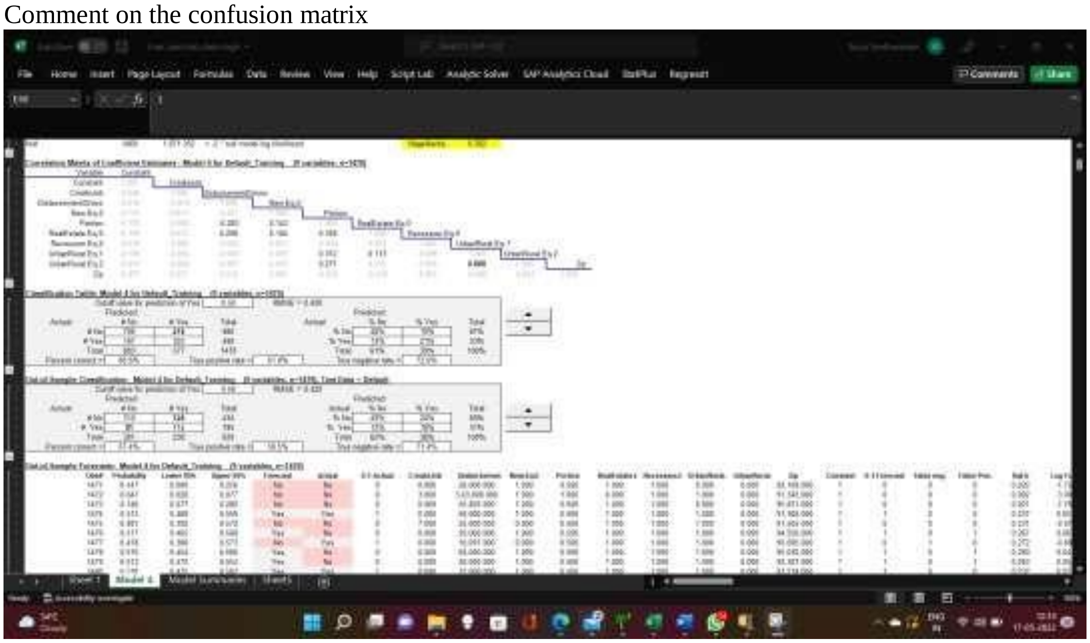
The new business start-ups- considering the no new small business start-up keeping it as a base the +1 addition may increase the need for availing the loan to new small start-ups in the city concerning an already existing business

The recession positive note- with keeping no recession as the base the +1 addition may indicate that the recession in the country would influence the loan demand in the city that may increase the bank to sanction more loan

The jobless people- the employees with a job as a base could infer that with +1 increase could indicate the demand for loans with jobless people.

The disbursement gross value- the amount of disbursement +1 increase could simultaneously increase the rate of loan sanction by the bank.

Comment on the confusion matrix



The true positive rate that is the sensitivity value is 61.8% in training data and 59% in testing data with a cut-off of 0.50

The true negative rate is 72% for training data and 71.4% for testing data from which we could infer that the specificity rate that is the total negative rate of loan sanction by the bank is more.

The accuracy rate obtained in the confusion matrix states that the value is around 68% correct in training data and 67.4% correct in testing data.

To improve the loan sanction by the bank, it will be advisable for the bank to collect some more new data to predict the availing of the bank loan by the people in the country.

3. Fit a neural network model to classify "default." (Dataset: loan_sanction_data- nn.xlsx")

a. Discuss the changes that you made to the hyperparameters to fit the model

i. To set the hyperparameters the variable that needs to be considered are

1. Learning rate=low value rate it keeps repeating idea values

2. Hidden and hidden nodes- depends upon the number of inputs

Structure of neural networks-connection of interconnected neurons, the hyperparameter layers consist of input layers that take in the input and passes to the network

3. The input layer includes how many nodes/neurons= no. of input variable or single variable

4. The output variable or layer consists of predicting the churn yes or no concerning one-note representing the churn and the gradient descent going on

5. Hidden layer consists of

a. 1 or 2 business problems stated as deep learning

b. Overfitting if we choose more hidden layer memorization

**How many nodes= 2/3(number of input nodes+no of levels of output nodes)**

c. Neurons<2*no of nodes in the input layer

d. The epochs and iterations that include the input variable, the dummy variable and the continuous normalise

e. For the loan sanction data provided the no of hidden layers that need to be implemented that depends on the total no of input and concerning the level of output

f. Here the input is default and the output includes 31 variables

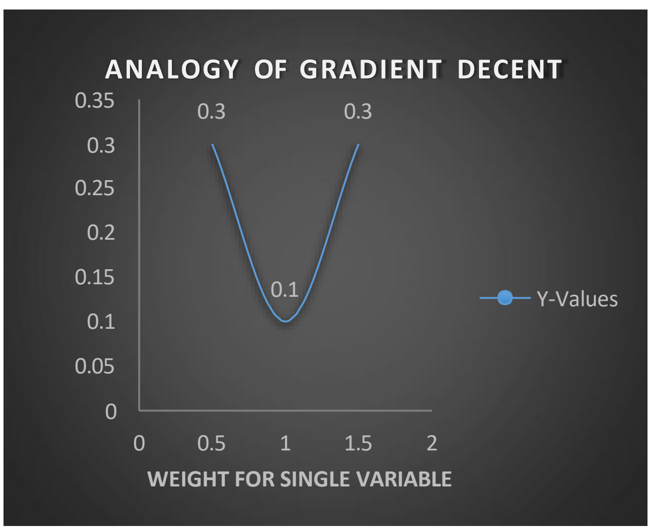
g. The summation equation includes

h. Z=w+w1x1+w2x2+w3x3+…wnxn

i. Includes the activation function that is G(Z): I/(I+EXP(Z)

b. Each neuron is the sum of the activation function that introduces non-linearity in the function

c. That leads to the calculation otherwise wise called the function approximation



MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. xx Issue x, Jan-June 20yy, Pages: 357-394

d.

e. The slope is negative as the x value increases the value of y decreases

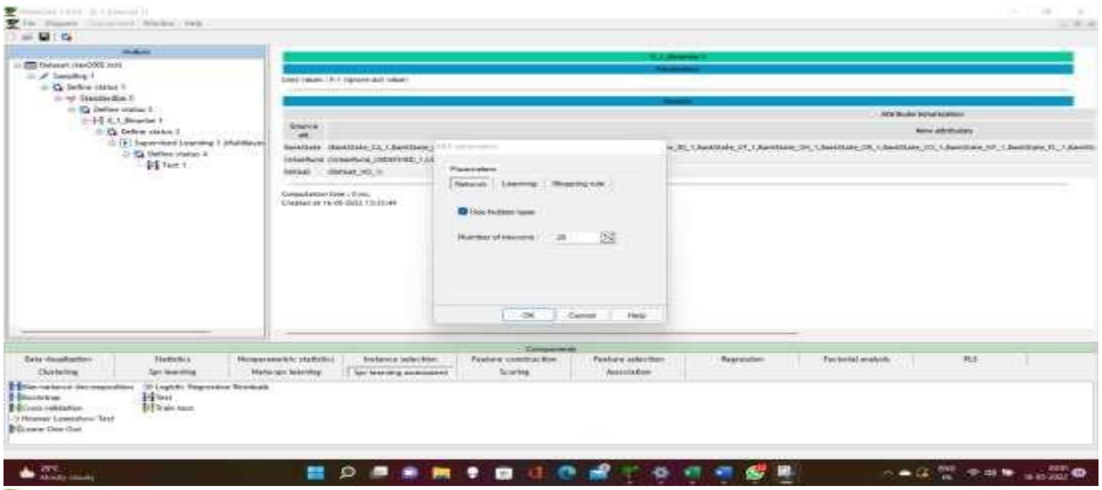
**f. How many nodes= 2/3(number of input nodes+no of levels of output nodes)**

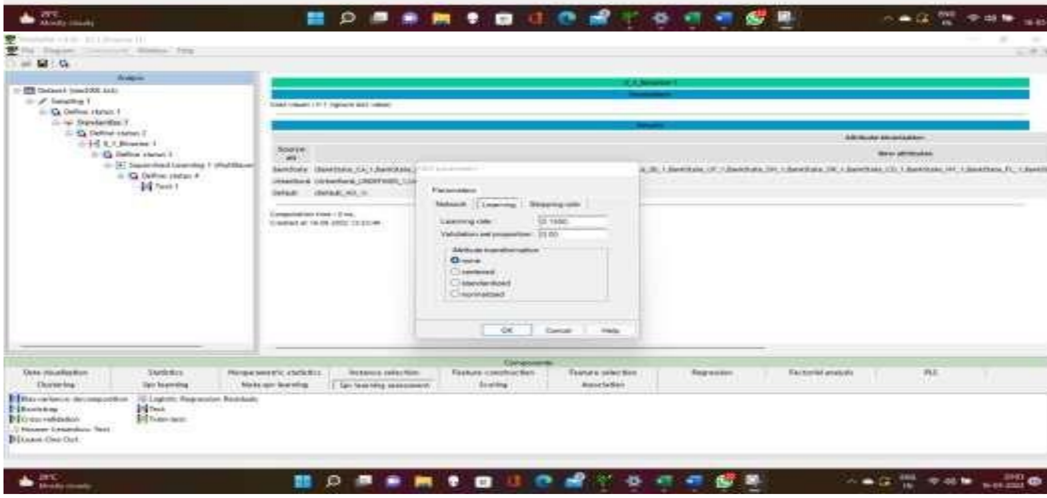
**i. =2/3(1+39)=2*13=26**

ii. That leads to the calculation of 26 nodes.

iii. The learning rate is 0.15

The validation set proportion is 0





b. How did you improve the predictive power of the model? Comment on the confusion matrix

i. THE PREDICTIVE POWER OF THE CONFUSION MATRIX has been improved by adding more variables like adding dummy variables that could relate to the three variables obtained in the confusion matrix that is specificity, accuracy and sensitivity

ii. THE MLP architecture training data output that we obtained infer the value FOR TRUE NEGATIVES AS 1 concerning FALSE NEGATIVE VALUE THAT GIVES THE VALUE OF 0.99.

iii. From which we could infer that the specificity rate seems to have a higher value than the sensitivity is a true positive rate

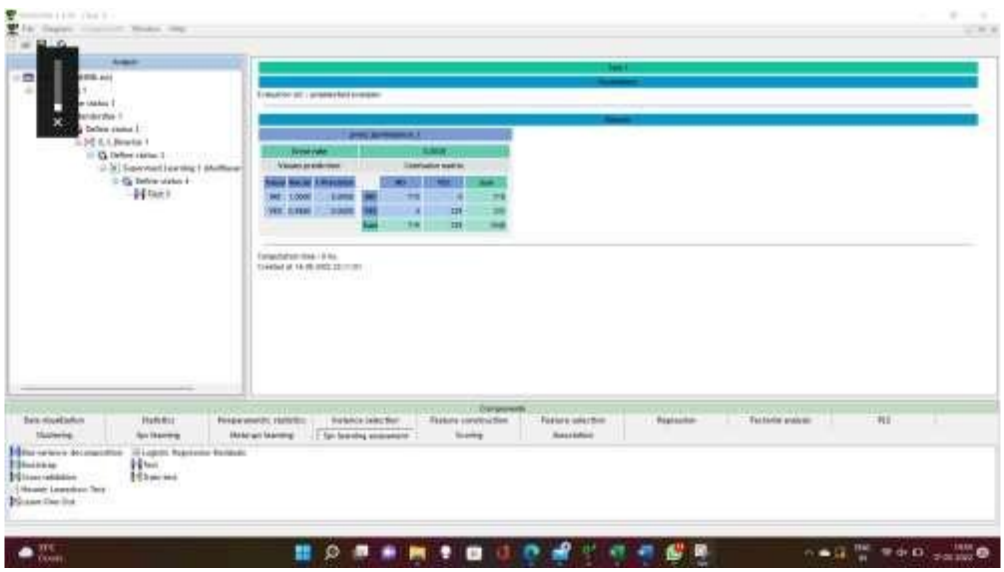
Hence it is been compared with the test data that has been inferred that the model is not fitting and both provide the same specificity rate which is the true negative rate

iv. Hence it is advised to the bank to avail some more data that may help in determining the more true positive

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. xx Issue x, Jan-June 20yy, Pages: 357-394

rate so that they can concentrate on sanctioning the loan more accurately.

c. 

4. The analyst is interested in identifying defaulters-. Which of the two models would you choose? Justify your answer

Logit regression analysis

- The predictive model mainly helps in identifying who is going to churn. It is mainly interested in predicting accuracy
- Model performs concerning the accuracy

To remove the insignificant value in the model

- P-value mainly explains about the sample matches the population, the goal of the predictive model is to build the training data concerning test if it works with the test data

- Three criteria with which the predictive model works is

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. xx Issue x, Jan-June 20yy, Pages: 357-394

a. Specificity
b. Accuracy
c. Sensitivity
d. Considering the cut-off value

- For each record model calculates the probability if greater than the cut-off value then the positive class or negative class can be classified
- If we check with training data and perform poorly with test it is called overfitting, memorized training data of out of sample.
- If the cut-off value is been adjusted in the confusion matrix then we can obtain a good fitting model to further determine the predictive nature of data using regression
- If the model is overfitting then use the stratified sampling, and adjust the cut off rate else try to gather data furthermore to make the model consistent and fitting.
- In this loan section data set, the data in the confusion matrix obtained is fitting and we decided to adopt this model as it has a more true negative rate with a 0.50 cut-off that is specificity
- To analyze the factors that are related to the output in the regression model are
  - VIF should not be greater than 5
  - The coefficient should be lesser than 0.05
  - P-value is pseudo R square value should be lesser than 0.198
  - The nagelkerke value should vary above 20%
  - The exponential coefficient are classified based on
    - Greater than 1 the loan is accepted
    - Equal to 1 the equal chances for sanctioning the loan and rejecting the loan
    - Lesser than 1 rejecting the loan
- NEURAL NETWORK ANALYSIS
- Learning rate=low value rate it keeps repeating idea values
- Hidden and hidden nodes- depends upon the number of inputs

Structure of neural networks-connection of interconnected neurons, the hyperparameter layers consist of input layers that take in the input and passes to the network

- The input layer includes how many nodes/neurons= no. of input variable or single variable
- The output variable or layer consists of predicting the churn yes or no concerning one-note representing the churn and the gradient descent going on
- Hidden layer consists of

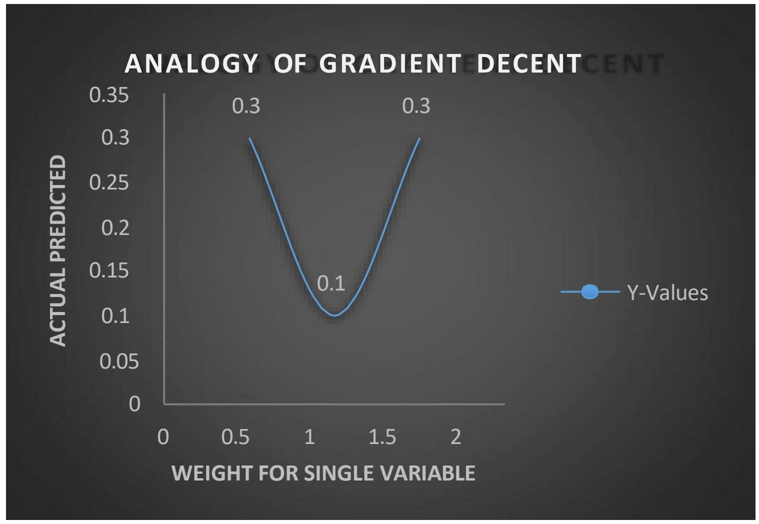
  - 1 or 2 business problems stated as deep learning

Overfitting if we choose more hidden layer memorization

  - **How many nodes= 2/3(number of input nodes+no of levels of output nodes)**
  - Neurons<2*no of nodes in the input layer
  - The epochs and iterations that include the input variable, the dummy variable and the continuous normalise
  - For the loan, sanction data provided the no of hidden layers that need to be implemented that depends on the total no of input and concerning the level of output
  - Here the input is default and the output includes 31 variables
  - The summation equation includes
  - Z=w+w1x1+w2x2+w3x3+…next
  - Includes the activation function that is G(Z): I/(I+EXP(Z)

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. xx Issue x, Jan-June 20yy, Pages: 357-394

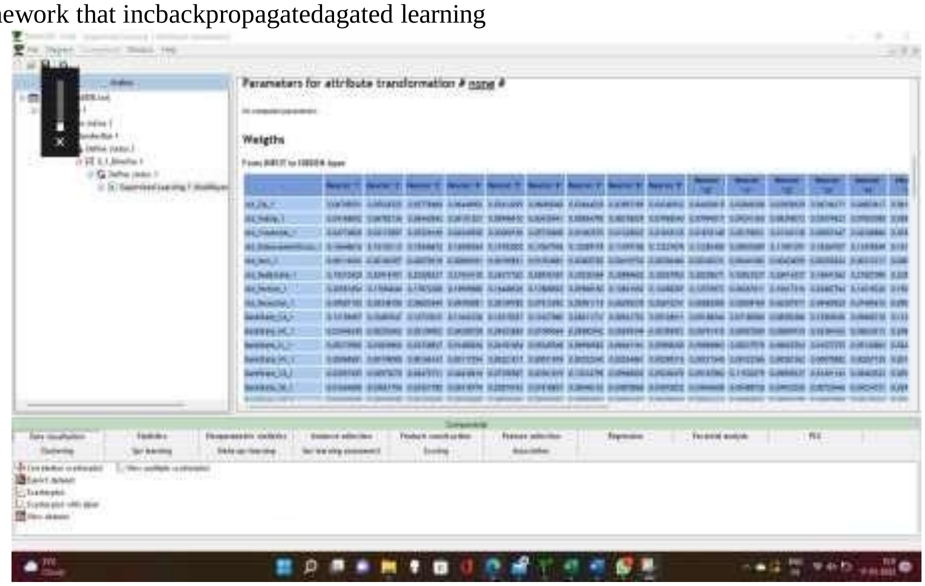
- 
- Each neuron is the sum of the activation function that introduces non-linearity in the function
- That leads to the calculation otherwise wise called the function approximation




- 
- The slope is negative as the x value increases and the value of y decreases
- **How many nodes= 2/3(number of input nodes+no of levels of output nodes)**
  - **=2/3(1+39)=2*13=26**
    - That leads to the calculation of 26 nodes.
    - The learning rate is 0.15
    - The validation set proportion is 0 SOME OF THE DISADVANTAGES OF NN ARE

- BLACK BOX -No of coefficient, weights will be output and may have different model
- Time-consuming it takes a long time to train the neural network products

The common framework that incbackpropagatedagated learning



- Analyzingng the output of both models

| ○ mode l name | ○ regressio n | ○ neural network |
|---|---|---|
| ○ 1 | ○ the predictive model mainly helps in identifying who is going to churn. it is mainly interested in predicting accuracy | structure of neural networks- connection of interconnected neurons, the hyperparameter layers consist of input layers that take in the input and passes to the network |
| ○ 2 | ○ it predicts the output data | ○ it works in terms of how the brain works |
| ○ 3 | ○ the input data can be compared with the test data to predict the future in terms of the forecasting model | ○ the error rate can be minimized by implementing the hyperparameters that include the hidden layer |
| ○ 4 | ○ the equation of regression is y=mx+c+…. | z=w+w1x1+w2x2+w3x3+…wnx n. includes the activation function that is g(z): i/(i+exp(z) o |
| ○ 5 | to analyze the factors that are related to the output in the regression model are vif should not be greater than 5 the coefficient should be lesser than 0.05 | the input layer includes how many nodes/neurons= no. of input variable or single variable the output variable or layer consists of predicting the churn yes or no concerning one note representing the churn and the gradient descent going on |
| | p-value is pseudo r square value should be lesser than 0.198 the nagelkerke value should vary above 20% the exponential coefficient are classified based on greater than 1 the loan is accepted equal to 1 the equal chances for sanctioning the loan and rejecting the loan lesser than 1 rejecting the loan | |
| | o | |

CONSIDERING THE ABOVE FACTORS OF CONDITION THE LOAN SANCTION DATA HAD BEEN MADE TO RUN THROUGH BOTH THE MODELS, AND THE FINAL OUTPUT THAT WE OBTAIN FROM THIS DATA ANALYSIS INCLUDES

The detailed explanation of the variables related to the data in the regression model is more accurate.

- The predictive model mainly helps in identifying who is going to churn. It is mainly interested in predicting accuracy
- Model performs concerning the accuracy
- To remove the insignificant value in the model
- P-value mainly explains about the sample matches the population, the goal of the predictive model is to build the training data concerning test if it works with the test data
- Three criteria with which the predictive model works is

a. Specificity
b. Accuracy
c. Sensitivity
d. Considering the cut-off value

- For each record model calculates the probability if greater than the cut-off value then the positive class or negative class can be classified
- If we check with training data and perform poorly with test it is called overfitting, memorized training data of out of sample.
- If the cut-off value is been adjusted in the confusion matrix then we can obtain a good fitting model to further determine the predictive nature of data using regression

If the model is overfitting then use the stratified sampling, and adjust the cut off rate else try to gather data furthermore to make the model consistent and fitting

- In this loan section data set, the data in the confusion matrix obtained is fitting and we decided to adopt this model as it has a more true negative rate with a 0.50 cut-off that is specificity
- And we could infer from the confusion matrix that the true negative that is specificity rate was higher, as well as the loan sanction, is recommended to collect some more data for the further accuracy rate
- When we compare it with a neural network we could infer that it can predict the error rate well and can implement the distance between the nodes and compatibility between the clusters of that particular neuron the similarity between the variables has been defined and the interest of variation within each variable has been stated to interpret the input value, the output value and includes the summation of weights and the activation function also.

Where both the test as training data fit hence this neural network can be accepted and it was fitted hence I would suggest **NEURAL NETWORK DEFAULTERS ARE BETTER COMPARED WITH REGRESSION MODEL**

## BIBLIOGRAPHY

### REFERENCES (APA Style)


1. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4), 589–609. https://doi.org/10.2307/2978933
2. Anderson, R. (2007). The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation. Oxford University Press.
3. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 54(6), 627–635. https://doi.org/10.1057/palgrave.jors.2601545
4. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
5. Breiman, L. (2001). Statistical modeling: The two cultures. Statistical Science, 16(3), 199–231. https://doi.org/10.1214/ss/1009213726
6. Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033
7. Chen, N., Ribeiro, B., Vieira, A., & Chen, A. (2013). Logistic regression and neural network classification of credit risk. Neural Computing and Applications, 22(5), 913–923. https://doi.org/10.1007/s00521-012-0851-4
8. Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964
9. Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. European Journal of Operational Research, 183(3), 1447–1465. https://doi.org/10.1016/j.ejor.2006.09.100
10. Fan, J., & Yao, Q. (2003). Nonlinear time series: Nonparametric and parametric methods. Springer.
11. Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics (5th ed.). McGraw-Hill.
12. Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. Journal of the Royal Statistical Society: Series A, 160(3), 523–541. https://doi.org/10.1111/j.1467-985X.1997.00078.x
13. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
14. Heaton, J. (2008). Introduction to neural networks with Java (2nd ed.). Heaton Research.
15. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). Wiley.
16. Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. Computer, 29(3), 31–44. https://doi.org/10.1109/2.485891
17. King, G., & Zeng, L. (2001). Logistic regression in rare events data. Political Analysis, 9(2), 137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868
18. Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001
19. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the International Joint Conference on Artificial Intelligence, 1137–1145.
20. Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proceedings of ICML, 282–289.
21. Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring. European Journal of Operational Research, 247(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030
22. Mitchell, T. M. (1997). Machine learning. McGraw-Hill.
23. Minsky, M., & Papert, S. (1969). Perceptrons. MIT Press.
24. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review. Decision Support Systems, 50(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006
25. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251
26. Ripley, B. D. (1996). Pattern recognition and neural networks. Cambridge University Press.
27. Thomas, L. C. (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. International Journal of Forecasting, 16(2), 149–172. https://doi.org/10.1016/S0169-2070(00)00034-0
28. Vapnik, V. N. (1998). Statistical learning theory. Wiley.
29. West, D. (2000). Neural network credit scoring models. Computers & Operations Research, 27(11–12), 1131–1152. https://doi.org/10.1016/S0305-0548(99)00149-5
30. Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7