E-147 Original

Evaluation of Neural Network and Logit models for classification of Default in a Honduran Bank

Walter Jeremías López Flores[1]

(1) Facultad de Postgrado, Universidad Tecnológica Centroamericana, UNITEC, San Pedro Sula, Honduras

INTRODUCTION

Banks have a wealth of data about their borrowers that can be used to predict the likelihood that borrowers will default on their loans. The evaluation of this type of risk is based on the probability that the borrower fails to comply with its obligations, a situation that in financial jargon is called default. Data mining is a promising area of data analysis that aims to extract useful knowledge from a huge amount of complex data sets, using statistical methods, especially Machine Learning (ML) techniques to model and predict losses from loan default.

OBJECTIVE

The purpose of the study was to evaluate the performance of Artificial Neural Networks (ANN) as modern techniques to classify the risk of default against the traditional Logit statistical method using a Honduran bank as a case study.

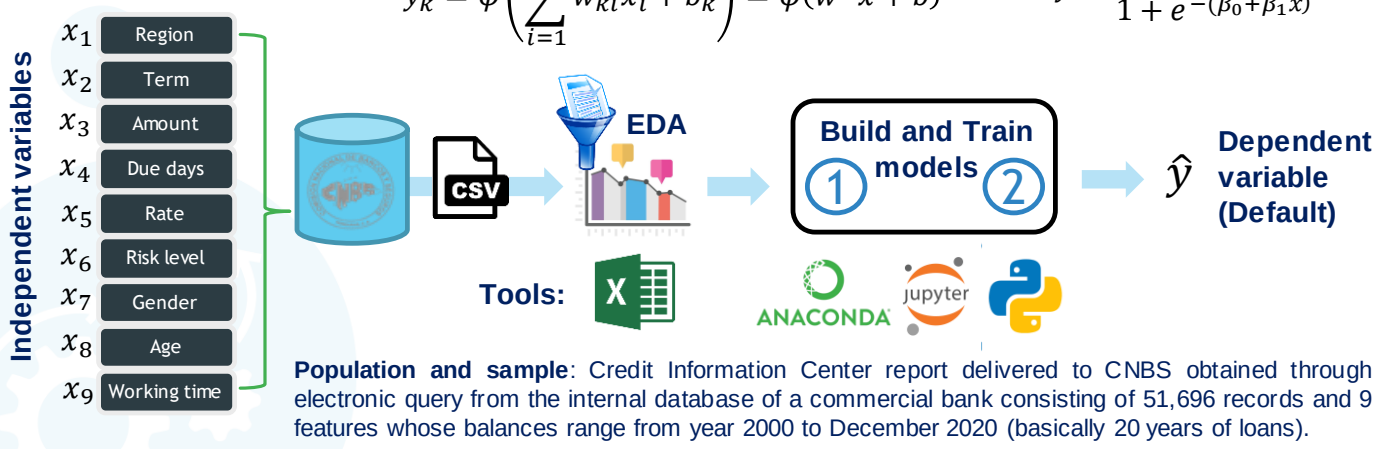
METHODS

Credit scoring model → ① **Multi-Layer Perceptron ANN** ② **Logistic Regression (Logit)**

$$f(x_1, x_2, \ldots x_m) = y_n$$

$$y_k = \varphi\left(\sum_{i=1}^{m} w_{ki}x_i + b_k\right) = \varphi(w^T x + b)$$

$$\hat{y} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



Population and sample: Credit Information Center report delivered to CNBS obtained through electronic query from the internal database of a commercial bank consisting of 51,696 records and 9 features whose balances range from year 2000 to December 2020 (basically 20 years of loans).

RESULTS

Table 1. Summary of results from Logit models

Resulting model params	1: base model	2: base + risk level	3: base + due days
Pseudo-R^2	-0.031	0.230	0.888
AIC	18,331.1178	13,692.0561	2,010.186
BIC	18,388.4671	13,757.5982	2,075.728
Log-Likelihood	-9,158.6	-6,838.0	-997.09
Coefficients (p-value)			
Term	-1.0197 (0.0000)	-0.7542 (0.0000)	-3.3273 (0.0000)
Amount	3.2866 (0.0000)	-1.3564 (0.2587)	2.3767 (0.4810)
Rate	-3.9312 (0.0000)	-4.9894 (0.0000)	-13.4547 (0.0000)
Working time	-2.2249 (0.0000)	-2.2045 (0.0000)	-3.4216 (0.0000)
Age	-0.7501 (0.0000)	-1.3624 (0.0000)	-1.1699 (0.0015)
Region	0.4539 (0.0000)	-0.3652 (0.0000)	-0.7861 (0.0000)
Gender	-0.3038 (0.0000)	0.3206 (0.0000)	-0.3458 (0.0058)
Risk level		5.2807 (0.0000)	
Due days			360.4526 (0.0000)

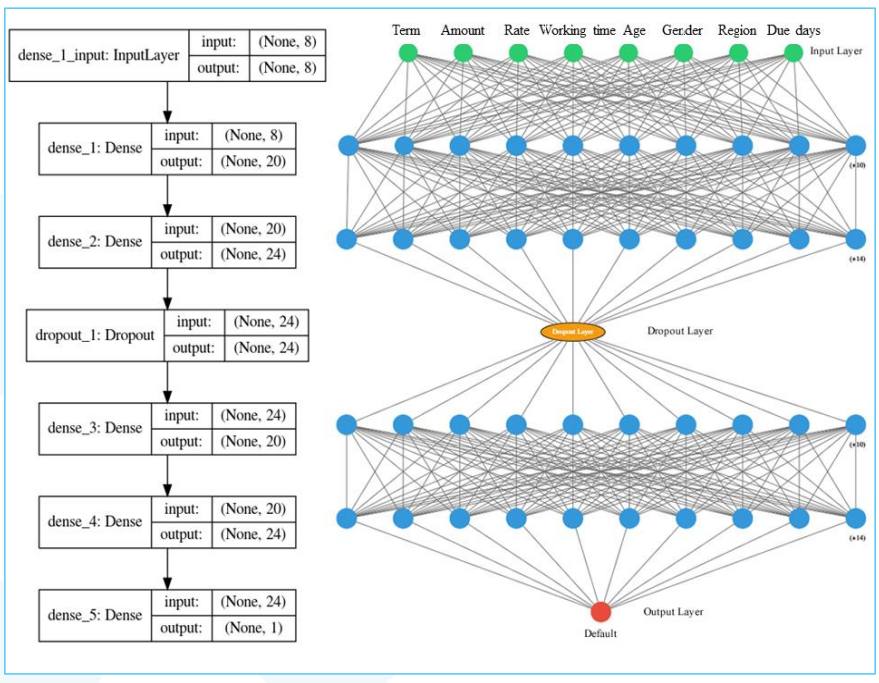


Figure 1. Diagram of the selected MLP model

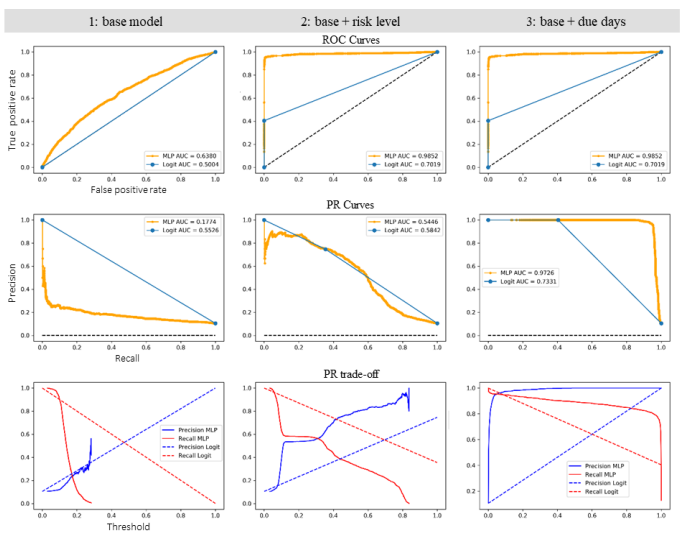


Figure 2. ROC and PR curves before SMOTE oversampling

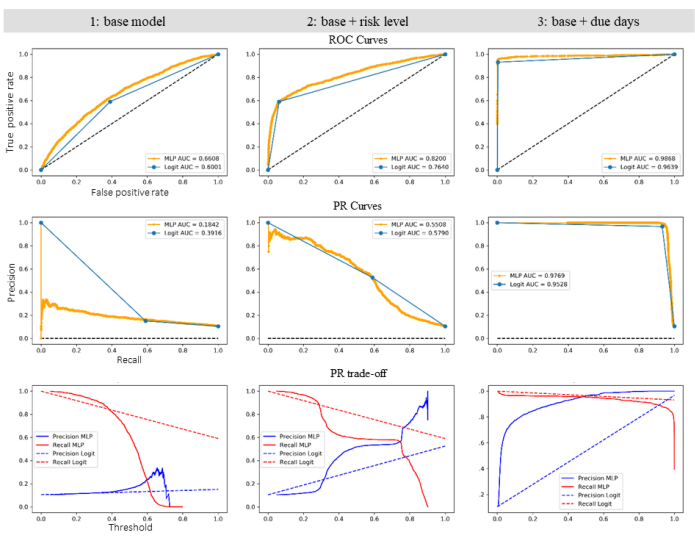


Figure 3. ROC and PR curves after SMOTE oversampling

CONCLUSIONS

The ANN outperformed the Logit method when the data is unbalanced. However, it was observed that when SMOTE oversampling was applied it slightly improves the performance of the MLP networks, but the Logit method indeed performed substantially better, but didn't reach the MLP's metrics: accuracy of 99.16%, precision of 99.47%, sensitivity of 99.59%, specificity of 95.48 %, F1-score of 99.53% and ROC and PR curves with AUC of 98.68% and 97.69% respectively.

Contact: walter.lopez@unitec.edu.hn **Conflict of interest:** none

CONICIETI 2024 I Conferencia Internacional de Ciencia, Tecnología e Innovación

AÑO DE Creatividad e Innovación 2024 · Alianza CARA Central America Research Alliance · PULTE INSTITUTE FOR GLOBAL DEVELOPMENT · INVESTIGACIÓN UNITEC · ceutec de unitec · unitec