




Identification and Control of Credit Risk in Banks Utilizing New Supervisory Technologies with Neural Network Algorithm and Random Forest Algorithm


Maryam Mashrooti¹, Ali Mohammadi^{2*}, Mehdi Mohammadi²

¹ Ph.D. Student, Department of Accounting, Zanjan Branch, Islamic Azad University, Zanjan, Iran



² Assistant Professor, Department of Accounting, Zanjan Branch, Islamic Azad University, Zanjan, Iran

* Corresponding author email address: Ali.mohammadi@iauz.ac.ir

Editor

Leila Youzbashi
Department of sport science,
Faculty of Humanities, University
of Zanjan, Zanjan, Iran
l.youzbashi@znu.ac.ir

Reviewers

Reviewer 1: Mehdi Rostami
Department of Psychology and Counseling, KMAN Research Institute, Richmond
Hill, Ontario, Canada.
Email: dr.mrostami@kmanresce.ca
Reviewer 2: Yaghob Badriazarin
Associate Professor of Sport Sciences, Tabriz University, Tabriz, Iran.
Email: badriazarin@tbzmed.ac.ir

1. Round 1

1.1 Reviewer 1

Reviewer:

The opening sentence, “In many countries regulatory authorities have the power to intervene to prevent bank failures or to address issues such as fraud or ethical problems,” lacks a citation. Adding a credible source would strengthen the claim.

The discussion of deposit insurance and moral hazard (Beck et al., 2006) seems incomplete without a more detailed explanation of how moral hazard directly impacts credit risk. Consider expanding this section to clearly link these concepts to your study.

The sentence “The integration of regulations with technological rules necessitates regulatory technology (RegTech)” could benefit from more in-depth background on RegTech, as it is a key concept in your research. Providing examples of successful RegTech implementations would enhance clarity.

The standard deviation values appear too small for some variables (e.g., education and gender). Ensure that these values are accurate and match the described population.

The results of the Area Under the Curve (AUC) and accuracy could be presented with confidence intervals to give a clearer picture of the variability of the model’s performance.

It’s stated that “the classification accuracy in the fourth cluster is higher than in the first, second, and third clusters,” but there is no explanation for why this is the case. Consider providing hypotheses for why certain clusters perform better.

Author revised the manuscript and uploaded the updated document.

1.2 Reviewer 2

Reviewer:

In the sentence “After collecting the previous bank customer data from the relevant database and cleaning the data,” clarify the exact methods used for data cleaning. Were specific statistical tests used to ensure data quality?

The formula presented for sample size calculation is valid, but more explanation is needed. Why was a confidence level of 98% chosen, and what specific assumptions were made regarding the standard deviation?

The feature selection process is mentioned but not thoroughly explained. Which specific algorithms were used for feature selection, and how were they compared?

The confusion matrix provides valuable insight into model performance, but the description of the findings could be enhanced by explaining how the accuracy of each model was measured and why the neural network performed worse in some clusters.

In Figure 1, there’s a reference to the Gain index but no detailed explanation of how this index is calculated. Including a brief explanation would help readers unfamiliar with this method.

In the paragraph stating “the random forest algorithm was more efficient in predicting customer credit risk,” consider providing a more detailed comparison between the performance of the two models. Why did Random Forest outperform Neural Network in this context?

The sentence “Banks in order to meet requirements and provide services... need to accurately identify them” seems somewhat vague. Can you specify which “requirements” and what type of identification methods you are referring to?

The historical account of credit scoring models is informative, but it would be useful to contextualize this within modern banking practices. How are these models being used with contemporary big data and machine learning tools?

The statement “The random forest (DT) used in this study had the highest accuracy among all models” could benefit from a brief explanation of how this finding aligns with existing literature on the use of decision trees in credit risk management.

Author revised the manuscript and uploaded the updated document.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.