



PREVENTING INTERNAL FRAUD IN MICROLENDING BUSINESS PROCESSES WITH MACHINE LEARNING MODELS: CONFIRMATORY FACTOR ANALYSIS (CFA) AND EXTREME GRADIENT BOOSTING (XGBOOST)

Heri Supriyadi *

School of Business
IPB University
Indonesia
heri_supriyadi@bri.co.id

Dominicus Savio Priyarsono

Department of Economics
IPB University
Indonesia
priyarsono@yahoo.com

Noer Azam Achsani

School of Business
IPB University
Indonesia
achsani@yahoo.com

Trias Andati

School of Business
IPB University
Indonesia
trias_andati@yahoo.com

*Corresponding Author email: heri_supriyadi@bri.co.id

Submission: 10 April 2021

Revised: 02 May 2021

Accepted: 19 May 2021

Peer-review under responsibility of 5th ASIA International Multidisciplinary Conference 2021 (Online)

Scientific Committee

<http://connectingasia.org/scientific-committee/>

© 2021 Published by Readers Insight Publisher,

Office # 6, First Floor, A & K Plaza, Near D Watson, F-10 Markaz, Islamabad, Pakistan,

editor@readersinsight.net

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



ABSTRACT

Internal fraud (occupational fraud) in the microcredit business has caused significant losses for the banking industry and financial institutions. The unique microcredit business process and the amount of loans provided were relatively small for each customer. The more extensive the credit portfolio in a bank /financial institution, it took many marketing personnel/microcredit analysts. Internal fraud was one type of operational risk that banks/financial institutions often face that focused on microcredit services. The most common types of fraud were Corruption and Asset Abuse (ACFE), such as Tempilan credit (loan partly used by the debtor), Topengan credit (misused loan), and Fictitious Credit. Machine learning that was run automatically was used to predict internal fraud in microcredit business processes. This study conducts the use of analysis with CFA (Confirmatory Factor Analysis) method to ascertain which component in the model of Social Identity and Triangle-Fraud most dominantly affects a person to commit fraud in the process of microloan services. This study also applies Extreme Gradient Boosting (XGBoost) model to predict the possibility of fraud events. The level of possible fraud events will be manifested in the form of "Risk-Scoring.". This research result is expected to provide input to the banking industry/financial institutions that serve microcredit to make efforts and make fraud prevention strategies more effective.

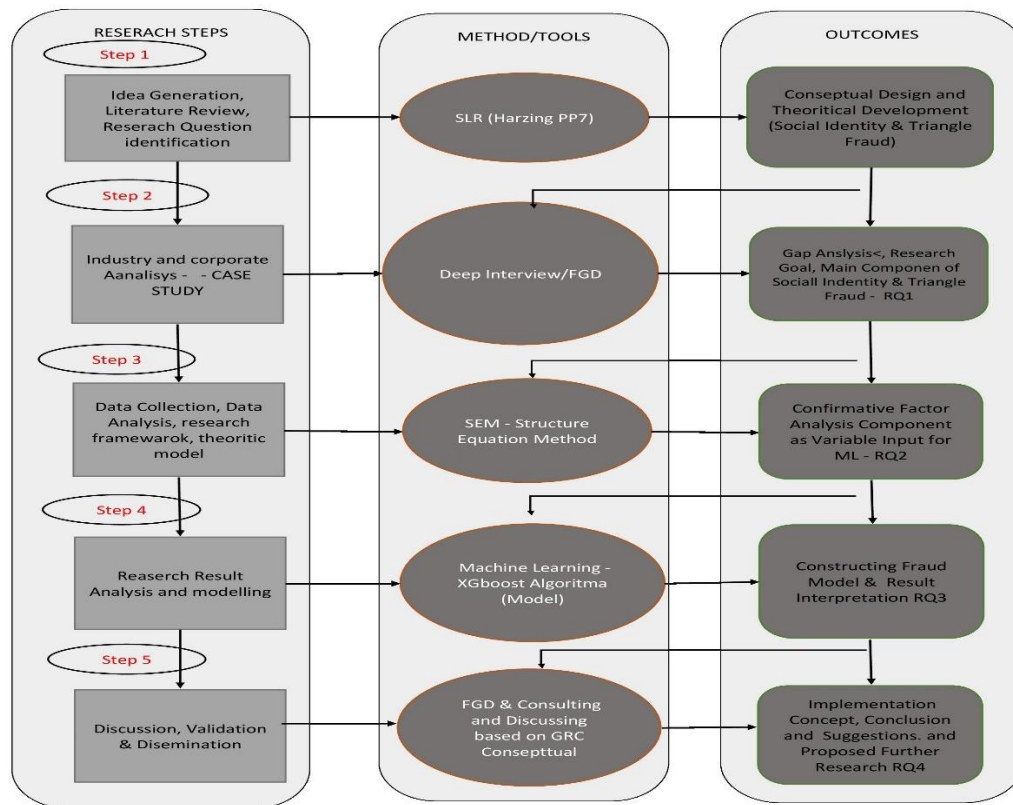
Keywords: *Fraudulent Models, Fraud Detection, Micro Credit, Loan, Machine Learning*

RESEARCH HIGHLIGHTS

The fraud is increasingly sophisticated, which results in the cost of company losses increasing from year to year. The banking industry is faced with a very crucial problem. Losses due to fraud are very complex, both financial losses due to eroding company profits and non-financial losses such as reputation losses. Non-financial losses include damage to reputation, loss of market position, negative investor sentiment, employee morale, and lost opportunities in the future. In 2019 according to the Association of Certified Fraud Examiners (ACFE, 2017) Indonesia, the most significant fraud was corruption, followed by misuse of company/state assets and assets.

Today, many large companies have invested heavily in new tools and techniques, including big data as the basis for artificial intelligence (AI) - a technology that is increasingly common today. This prediction system is designed primarily to assess better risks in applying for loans to banks, especially those caused by internal company factors.

GRAPHICAL ABSTRACT



Research Objectives

To understand the occurrence of fraud, this study examines the behavior of "Relationship managers" on their work and the strategies that bank "X" has in reducing the occurrence of fraud. Ultimately, this study aims to analyze and confirm as follows: 1) Determine the main components of social identity which are used as input variables to determine the causes of fraud, 2) To find the primary/dominant cause of fraud in the microloan process at microfinance institutions based on the Fraud Triangle theory, and 3) Finding a red flags-based predictive analytics model to predict the occurrence of fraud risk in the microloan business process. With the explanation above, this research was expected to provide the following benefits: (1) This research can be a reference for decision-makers in the banking industry and micro and small financial institutions in making operational cost savings, especially the prevention of losses due to fraud, (2) Providing alternative solutions for the detection of compliance violations and worker integrity (3) Creating a culture in the use of Machine Learning-based technology innovations in the banking industry and micro-loan institutions to create more efficient business processes.

Methodology

In this study, the authors used a mixed quantitative and qualitative approach. The quantitative analysis describes and measures the amount of influence between the

dependent variable and the independent variable. The researcher used operational variables, namely the dependent variable on fraud, and the second, namely the independent variable, divided into variable opportunity, rationalization, and pressure. The author's primary data was data on fraud incidents during 2017, 2018, and 2019. The fraud data includes fictitious loans, temporary loans and mask loans, gratuities, and delays in debtor installments. Secondary data were taken to support data relevant to fraud incidence in microloan service activities at banks / other micro-financial institutions, sourced from industry, associations, and other relevant institutions. The initial questionnaire was used to test its validity and reality, followed by a revision of the final questionnaire. Before revising the questionnaire, the authors conducted a series of surveys and interviews to capture the field's realities and link them and match them with the Fraud Triangle theory.

Results

Results of first research step conducted by deep interview and focus group discussion include components of social identity and triangle fraud theory. The components then are analysed by *Structure Equation Model (SEM)*, *Confirmatory Factor Analysis (CFA)* to identify which the most significant components of social identity and triangle fraud theory influence occupational fraud. The components will be used as variable inputs for machine learning to construct the predictive analytic model to predict the possibility of fraud. Value of the prediction illustrated as risk scoring. In this step, interpretation of the risk scoring is also discussed.

Finally, the research will also propose the implementation strategies of the risk scoring as a integrated part of fraud risk mitigation based on the Governance, Risk and Compliance framework in the company.

Findings

Modelling Employee Risk Scoring (ERS) for marketing personnels/ credit analyst ERS will be done using XGBoost (Extreme Gradient Boosting) method, as a Classification Technique in machine learning, and will be illustrated in the ensembles decision tree. The reason for the selection of XGBoost method is because this method is a method that has high accuracy with low computing power. Employee Risk Scoring (ERS) & Machine Learning (ML) modeling is a Classification Technique in machine learning and can be integrated into an ensembles decision tree. ERS will be reference for supervisors to manage the people and to mitigate occupational fraud risk in the business process. It is as a part of integrated implementation strategy based on GRC framework that will be designed to mitigate the fraud risk in the bank

References

ACFE Indonesia. (2017). *Survey Fraud Indonesia 2016*. Jakarta: ACFE Indonesia Chapter.