

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

Heri Supriyadi, IPB University  
Priyarsono, D.S., IPB University  
Noer Azam Achsani, IPB University  
Trias Andati, IPB University

## Abstract

*Internal fraud in microcredit business processes has caused significant losses for the banking. Internal fraud was one type of operational risk that banks/financial institutions often face that focuses 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 is run automatically was used to predict internal fraud in microloan business processes. This study apply Extreme Gradient Boosting (XGBoost) model to predict the possibility of fraud events. The level of possible fraud events was manifested in the form of "Risk-Scoring." This study also conducts the use of analysis with CFA (Confirmatory Factor Analysis) method affects a person to commit fraud in the process of microloan services. This research result is expected to provide input to the banking industry that serve microcredit to make efforts and make fraud prevention strategies more effective.*

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

## INTRODUCTION

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. Thus, the authors see how important to see the readiness of the banking industry in implementing effective measures to combat fraud. 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) Indonesia, the most significant fraud was corruption, followed by misuse of company/state assets and assets.

Hence, what should companies do to identify and prevent fraud? What are the gaps/challenges faced by the company in improving the implementation of risk management and strengthening internal control to reduce the incidence of fraud? Many companies in the banking industry have special programs to reduce the risk of fraud. However, most companies only have policies, procedures, and training resources in implementing risk management. Companies must implement effective risk management, including utilizing internal input from various company stakeholders in reducing the risk of fraud.

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. Such technology makes it easier for companies to make decisions about predicting and preventing fraud more accurately. With this prediction tool, the company's focus is more on aspects of social behavior, including the transaction behavior of employees or customers. This prediction system is designed primarily to assess better risks in applying for loans to banks, especially those caused by internal company factors.

In empirical studies, currently, many researchers have linked the incidence of fraud with the social identity it bears. According to social identity theory, some individuals who think that 'honesty and shame' are essential can behave more praiseworthy and ethically than individuals in the group who think otherwise. This social identity is significant for the banking industry because the banking industry is closely related to the concept of business or money. Meanwhile, industries closely related to the concept of money are always hypothesized as having a 'dishonest environment.' The research found that the social identity component is included in the fraud indicator component (Red flags). The prevailing fraud indicators are limited to the 'Triangle fraud' theory - Cressey (1953); Diamond fraud - Wolfe & Hermanson (2004); Pentagon fraud - Marks (2009); 3C - Razee & Riley (2010) etc.

In the fraud triangle theory, fraud perpetrators can always take advantage of every opportunity due to the weaknesses of every control in the company (Padgett, 2015). Therefore, an appropriate and effective strategy to reduce the incidence of fraud must be made, among others, through institutional improvement, strengthening internal control, implementing risk management, strengthening audit quality, etc. For that, several problems can be identified as follows: 1) what are the main components of social identity that can be used as input variables to determine the causes of fraud? 2) Which components of the Triangle Fraud theory, *i.e.*, Opportunity, Rationalization, and Pressure, have the most influence on the possibility of fraud? 3) Can the red flags-based predictive analytics model be used to predict the occurrence of fraud risk in the microloan business process? And 4) how is the implementation of the predictive analytics model using ML in mitigating the risk of fraud in the microloan service process at banks?

The use of technology is currently being carried out to increase fraud prevention capabilities; Baz et al., (2016) stated that fraud is a fundamental problem in every financial institution. Technological advances have led banks to develop electronic banking service technology, which has also created new challenges in preventing fraud in banks. The results of his research have succeeded in developing a fraud prevention framework with administrative and technological controls. Hence, this conceptual research develops on predictive analytics to build analytical models that predict the likelihood of fraud as written in the research objectives in the previous chapter. Therefore, the research will cover a series of methodologies that can manage various big data-based problems that are starting to be managed and utilized by various industries, including the banking industry. It is an iterative process that links several statistical methods from sampling, model estimation, model prediction, and evaluation to form an integrated system that can predict fraud, violations, and other things that can harm the company.

## LITERATURE REVIEW

Several main theories regarding fraud showed that several contributing factors can increase fraud. Bureaucracy and inefficient administrative and political structures were the main contributing factors that increase the level of corruption (Goel & Nelson, 2010), wage levels (Van Rijckeghem & Weder, 1997), levels of urbanization, and globalization (Goel & Nelson, 2010). However, several factors were also found to reduce corruption (Bhattacharyya & Hodler, 2015), including civic

participation/press freedom and ethnic diversity. Other empirical studies also assessed the factors that can reduce the causes of corruption were ethnic diversity (Treisman, 2000) and the flow of globalization. They could also reduce corruption, especially for developing countries (Badinger & Nindl, 2014); High economic growth can also effectively reduce corruption (Bai, et al., 2013; Paldam & Gundlach, 2008). Another empirical study found that higher levels of education correlate with lower levels of corruption (Glaeser & Saks, 2006). Meanwhile, increased use and availability of the internet also correlates with lower levels of corruption (Andersen, et al., 2011). Microdata can also be seen as an attempt to measure corruption (Chatterjee & Ray, 2012).

Alain Cohn, Ernst Fehr & Michel Andre´ Mare-chal (2014) found that, on average, bank employees behave good-manner only when there was reasonable control. Alain Cohn et al., (2014), inspired by George Akerlof & Rachel Kranton, (2000, 2002, 2005), states that individuals have several social identities based on gender, ethnicity, or profession. Identity is associated with specific social norms. Identity and norms are related and relevant, depending on the individual's attributes of their identity. Alain, Cohn et al., Suggest that banks should encourage honest behavior by changing norms related to the professional identity of their workers. The banking industry is intrinsically linked to the concept of money, and that the condition of a professional identity making the concept of money very important. The concept of money triggers selfish behavior; a condition of professional identity can also lead to more selfish behavior, namely dishonesty. For this reason, management must push their professional identity to society. Probably, it is requiring bank employees to take the professional oath; such oaths encourage bank employees to consider the impact of their behavior on society rather than their short-term benefits. This change in norms may be an essential step towards promoting sustainable business culture change.

From a theoretical perspective, the three factors in the triangle theory of fraud were opportunity, motive/incentive, and rationalization. The three elements depend on each other. However, the complex personal and ideological factors associated with fraud motivation and rationalization of fraud (Albrecht, 2012) has been evaluated less by researchers than the opportunity factor. It is because of the factors that motivate people to commit fraud, and it is more difficult to predict how different people will respond to opportunities for fraud in the workplace. Clarke & Felson (2004); Lyman & Potter, (2007) applied a rational choice perspective to explain fraud in lower-level employees. Financial stress factors - be they personal, external, or work-related and incentives for financial gain are the main motivating factors found in low-level employee offenders Matthew Hollow (2014); Albrecht, (2012). Matthew Hollow (2014) said there was strong evidence to suggest that, in the banking and financial industry, the motives for committing fraud differ significantly between lower-level employees and upper-level employees. Lower-level employees tend to be motivated by personal pressure, while upper-level employees are motivated by external or work-related pressures. Incentives by lower-level bank employees tend to extract money from the organization, and vice versa for upper-level employees.

The same condition occurs in Indonesia, where there are still many banks that experience fraud incidents (Meliana, 2019). This fact shows that corruption can never be eliminated, but it can be reduced, and the opportunity for corruption to occur can be minimized. Azem, et al., (2017) conducted research related to governance and corruption in Bank Garments. A system is used to prevent and assist workers from a moral perspective, not to commit corruption. This study presents several general practices that apply in banks to minimize behavior that leads to corruption, including vital monitoring, decentralization of authority, review of decision-making processes, high audit intensity, disciplinary processes, and transparency and anti-corruption culture.

In the microfinance institution industry, Hartugi, (2007) has researched how microfinance institutions strive to minimize fraud. The existence of this microfinance institution cannot be separated from the implementation of good financial risk management. Thus, it can be concluded that microfinance institution's success is also determined by, among others, internal supervision, audit capabilities, and exemplary risk management implementation.

Empirical research related to fraud cases reveals that there were indications that can be seen before fraud occurs (Deloitte, 2012). This indication was called red flags (Mohamed & Jomitin, 2014). Red flags are specific indicators that indicate a potential occurrence of fraud. Red flags are conditions that indicate motivation and opportunities for potential fraud (SA 240). Red flags are an essential mechanism for detecting fraud early. The increasing number of frauds in financial corporations has raised concerns for corporate stakeholders. Therefore, financial corporations have started to consider Red flags as necessary information in making early warning tools for fraud. Signs of possible fraud or often called red flags are events or sets of conditions that can inform the organization about fraud (Padgett, 2015). Meanwhile, according to Stamler, et al., (2014), Red flags were defined as a structured process for sharpening, disclosing, and documenting fraud, suspected theft, and corruption so that they can be identified more quickly and efficiently. Specifically for conditions in Indonesia, Irianto & Novianti (2018) stated that early detection of fraud could be seen from the fraud symptoms that have occurred. These symptoms usually occur due to weak internal control systems, anomalies in accounting records, "oddities" in the behavioral aspects of fraudsters, and complaints. Thus, many other indicators can be used as red flags for the occurrence of fraud.

Huang et al., (2017) have sorted each component into several more specific indicator sections. They concluded that five essential parameters influence fraud: poor performance, external financial needs, financial pressure, lack of management supervision, and market competition. Red flags are not just signs of fraud. However, Red flags can be used as an essential component in preventing fraud. Many Red flags can be used as variables that affect the occurrence of events in the company. It is essential to observe and follow up on messages from the red flags that appear. Messages from red flags can be obtained in plain view or through analysis both empirically and descriptively.

Adequate policies and SOPs can control incidents of internal fraud and wrong actions, Nawawi & Salin, (2018). However, they also warned that policies and procedures would not be effective and would not function when the people who are supposed to control risks commit their violations of policies and procedures. The low level of compliance with internal controls provides an opportunity for fraud to occur, in line with the fraud triangle theory. In line with the above, Suh et al., (2019) concluded that the control mechanism associated with the measure of prevention tools (qualitative aspects of internal control) is more important than the existing anti-fraud control alone. They stated that management who ignored control had a higher risk than the incidence of collusion in the company. This research is also confirmed by Meliana's, (2019) findings in the banking industry in Indonesia. Some people think that their position and position can be overreacted, and the ineffective condition of company values is even used as a justification for acting deviant.

The most effective fraud prevention and detection procedures are operational audits, internal control reviews and improvements, cash reviews, code of sanctions for vendors, and ethics officers, Zamzami et al., (2016). While Lambe, (2013) adds that managing opportunities to prevent fraud is essential, strengthening internal control in several ways can narrow the opportunities for fraud to occur. However, internal control itself will not prevent fraud because 20 percent of fraud is carried out randomly, regardless of control problems, and 11% of fraud is committed because there is collusion to trick control.

## RESEARCH METHODS

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. This analytical approach quantitatively describes and measures variable target opportunity, rationalization, and pressure. The theoretical review of the analysis is referred to as part of the Literature Review. Simultaneously, the conceptual framework is a brief scheme of the main stages the researcher will take to make decisions. 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 qualitative approach is used to create a fraud risk mitigation strategy. This stage exploits information related to fraud that occurred at bank 'X,' a practice that has been done in several places, and compares these results.

The primary data used by the author is data on fraud incidents during the period 2017, 2018, and 2019. The fraud data includes fictitious loans, temporary loans and mask loans, gratuities, and delays in debtor installments. Secondary data are taken to support data relevant to the incidence of fraud 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 realities in the field and link them and match them with the Fraud Triangle theory. The questionnaire is standardized, in which respondents are asked the same questions in the same order, and the results are expected to be consistent. Meanwhile, a comparative analysis is used as the primary function to analyze data, line by line, to capture the concepts and relationships of all the variables used.

In the data collection stage, the first thing to do is conduct a series of semi-structured interviews with several stakeholders from the selected sample. Respondents are expected to have experience and responsibilities directly related to the microloan process at Bank 'X'. Samples and respondents were selected carefully to reflect the actual events. The author collects data, analyzes, and tests the fraud triangle theory to find results that are interrelated with one another. The second step is to collect secondary data from industry data, government data, and other published data. This data is essential for the writer to determine the main points that must be addressed during interviews and observations. The first and second steps are expected to make it easier for the author to identify key dimensions and modify the model or frame of mind established through the initial hypothesis, revise the initial questionnaire, and adapt the appropriate questionnaire.

This research is in the form of a "case study" on the Bank "X," and for research purposes, the author uses a "pseudo name." It is used to protect the confidentiality of an institution. The research method used is empirical. With this method, empirical evidence can prove the truth of an event. Thus, this study's primary data were taken from population data in Bank X's internal database, namely data of all marketing officers or Relationship Managers (RM). This data includes, among others, performance profile data—demographics, career trail, financial transactions, and financial liabilities data during the study period above. The sampling technique was carried out using purposive sampling. The sample in this study is a marketing officer or Relationship Manager (RM). Sampling location in Jakarta. The determination of the number of samples in this study refers to an approach based on a standard formula (Bartlett, Kotrlik & Higgins, 2001).

## DISCUSSION

The preliminary findings through the induction process are expected to add to the perspective of researchers in this study. They will add the ability to replicate the actual situation to confirm hypothetical cases or findings that are contradictory to the selected cases. The study population was marketing officers or Relationship Managers (RM), business and operational managers who had relevant responsibilities. This research is built on a conceptual theory by developing a structural model to see the variables that influence fraud. The author determines the variables involved based on the existing conceptual framework of the theory and then tests the model for the relationship between the variables.

Losses due to fraud in the loan process are a significant component of the bank fee structure. Therefore, efforts to increase efforts to prevent fraud by modeling fraud can be interpreted as austerity efforts. That is why ML technology has begun to be used frequently in efforts to prevent fraud. The loan process in banking begins with a submission process, analyzes it up to a decision, and ends with repayment. One of the tools used for the analysis process of a credit application is a credit score model (Credit Risk Scoring / Credit Risk Rating). In preparing the credit score, modeling is usually based on Logistic Regression (LR), *i.e.*, accepted or rejected. Like credit score modeling, modeling through ML must also go through training on historical data, primarily related to credit history data and relevant red flag data, to get accurate modeling. The ML model assesses the importance of each provided attribute and translates it into a prediction. The main limitation of such an ML model is that it can only consider the linear dependency between the input variable and the predicted variable. This modeling makes the logistic regression interpretable. This widespread logistic regression model has been used in credit risk modeling. ML allows the use of more sophisticated modeling techniques, such as introducing non-linearity to the model and detecting more complex dependent variables. This study will use the XGBoost model by first looking for the selected variables accurately through the CFA approach.

<b>Table 1</b>	
<b>THE RULE OF THUMB FOR THE REFLEXIVE INDICATOR CONSTRUCT USED</b>	
The goodness of Fit Indices	Cut-Off Value
X <sup>2</sup> Chi-Square	Diharapkan Kecil
Probability	≥ 0.05
CMIN (Chi-Square) / DF (Degree of Freedom)	≤ 2.00
RMSEA (The Root Mean Square Error of Approximation)	≤ 0.08
GFI (Goodness of FIT Index)	≥ 0.90
AGFI (Adjusted Goodness of FIT Index)	≥ 0.90
TLI (Tucker Lewis Index)	≥ 0.95
CFI (Comparative Fit Index)	≥ 0.95
RMSR (Root Mean Squares Residual)	≤ 0.05

CFA is part of Structural Equation Modeling (SEM) which helps test the validity and reliability of construct-forming indicators or latent variables of the 'fraud' triangle, so it is expected to explain: 1) Does the variable which consists of the three components of the Triangle Fraud theory have a relationship or influence each other on the incidence of fraud?, 2) Based on the validity of the measurement model, factor loading, chi-square test and good-fit models such as CMIN, RMSEA, GFI, AGFI, TLO, CFI & RMSR, which factors are the most dominant in forming the construct components of 'opportunity,' 'rationalization' and 'pressure' in the fraud Triangle

theory and 3) What combinations of variables influence each component of the Triangle Fraud theory to form construct fraud that is then used for prediction models by XGboost?

At the analysis stage, the Confirmative Factor Analysis component is used as an input component for Machine Learning (ML). The XGboost machine learning algorithm results are in the form of a predictive analytics model that can predict the occurrence of fraud risk in the microloan business process at the company that is the object of this dissertation. At this stage, it will be discussed in breakdown according to the steps using the XGboost algorithm. The prediction results from the resulting model are expected to provide a better alternative for the company to avoid fraud. The output produced at this stage is the predictive analytics model, the construct fraud model, and the results' interpretation.

CFA is a "Supervised" algorithm in ML used to reduce the selection of incompetent variables. This algorithm represents the general variance, namely the variance due to the correlation between the dependent variable and the variable independence on the fraud prediction model. CFA in the XGBoost model reduces many unnecessary variables to only a few of the most critical variables. The selected variable has several factors that contain several factor variances as a whole. For this reason, the selection of 'Eigenvalue,' which has a factor of  $> 1$ , is the key to choosing an accurate fraud prediction model in the XGBoost model. According to Bolton & Hand (2002), in Supervised Machine Learning, 'variables or data' both who commit fraud or do not commit fraud are used to create models. At the same time, Gradient Boosting generates feature importance to explain the influence of each independent parameter on dependent parameters. This method will generate some ruling trees with multiple nodes (leaves). The reasoning system is used to explain and describe the relationship between variables to be studied. Study variables, *i.e.*, independent variables expressed on  $X_1, X_2, \dots, X_n$ , while the dependent variable is fraud in fraud on micro-credit (Y).

Confirmatory Factor Analysis (CFA) is a multivariate statistical procedure used to test how well the measured variable represents several constructs. The researcher determines the required number of factors associated with the construction variable. The procedure carried out is; first, the researcher defines individual constructs. This procedure involves a pretest to evaluate construction items and a model measurement confirmation test performed using Confirmatory Factor Analysis (CFA). Second, the researcher develops a theory of the overall measurement model. In this Confirmatory Factor Analysis (CFA), the researcher considers the unidimensionality. The third is that researchers design studies to produce empirical results, in which the measurement model must be determined. Finally, the researcher assesses the validity of the measurement model, *i.e.*, assessing the validity of the measurement model occurs when a theoretical measurement model is compared with an accurate model to see how well the data fits. The validity of the measurement model is in the form of a loading factor which must be greater than 0.70, the chi-square test, and good-fit models such as RMR, GFI, NFI, RMSEA, SIC, BIC, etc.

The purpose of the factor analysis is to find the minimum possible factor with the principle of simplicity (parsimony), which can produce a relationship or correlation between the observed variables. The procedure taken calculates the correlation matrix to determine the adequacy requirements and looking for factors or Extracting Factors (extracting factors) that can explain the correlation between the indicators studied and performing factor rotations that can optimize the correlation between the independent variables observed. So, CFA can reduce the number of influential variables to be used in further analysis, namely through the Employee Risk Scoring (ERS) modeling to predict fraud.

Meantime, XGBoost was developed by Chen & Guestrin (2016) and the variant of the tree gradient boosting algorithm. This algorithm is an interpretation of Newton's method in space functions; Nielsen researchers (2016) call it Newton Boosting because of similarities in the concept

of the algorithm. Optimizations performed by the XGBoost algorithm are ten times faster than other gradient Boosting implementations (Chen & Guestrin, 2016). The Analysis and Modeling results will be used to create a fraud risk mitigation strategy based on the GRC framework. XGBoost algorithm can perform various functions such as regression, modeling, and ranking. XGBoost is a tree ensembles algorithm consisting of several CART collections (Classification & Regression Trees). The most crucial factor behind XGBoost's success is scalability in a variety of scenarios. This scalability is due to the optimization of the previous algorithm.

This success was evident when XGBoost became one of the methods that are being widely applied to various cases in ML. XGBoost was first introduced in the Higgs Boson Competition. Gradient boosting is a regression and classification algorithm that applies the ensemble concept of weak predictors and generally uses decision trees. Optimization is done using boosting techniques by optimizing the value of the loss function. The loss function is the evaluation mechanism of the model. The high loss function value means that the resulting model is inferior and vice versa. Gradient boosting combines iteratively weak predictors by minimizing the mean square error ( $\hat{y}-y$ ) of the model, where  $F$  is  $\hat{y}=f(x)$ . The process of each iteration produces a collection of hypotheses that form a model and produce a predictive value.

Gradient boosting performs optimizations in each iteration of  $m.1 \leq m \leq M$  by building an  $F$  model. Like the boosting method in general, residual value  $y- F(x)$  in the model is a negative gradient value. So, in short, gradient boosting is a gradient descent algorithm that reduces the loss value in a differentiable way so that it leads to a gradient. The end goal is to get the closest function  $F(x)$  to the building functions  $f(x)$  by minimizing the value of loss function  $L (y, F(x))$  with equations below:

$$F = \operatorname{argmin}_F \int E_x, y [L (y, (x))]$$

In the training process series, each iteration is created so that the average value of the loss function becomes minimal based on the original function  $F_0(x)$ . Generally, gradient boosting algorithms use the following Equations.

$$\{\gamma m h_m\} = \operatorname{argmin}_{\gamma} \sum L (y_i, (m-1) (x_i) + \gamma m h_m(x_i))$$

This study uses thirty-two variables and the method used by Shaio Yan Huang, Chi-Chen Lin, An-An Chiu & David (2016), where they use the Lawshe approach as a variable to predict fraud. Use of the Lawshe approach to justify the content validity of each factor via CFA to ensure representativeness. While the remaining thirteen are unique variables that are domestic identity. The researcher also used one operational dependent variable on fraud and several independent variables classified into the elements of Opportunity, Rationalization, and Pressure. The dependent variable is used to explain how fraud can occur. According to the identity theory discussed earlier, the independent variable is used to predict the identity conditions and behavior of marketers who commit fraud. Attached in Table is only a part of the operational variables used in the study, where the CFA determines the final operational variable.

## CONCLUSION

ERS modeling in this study will be done using XGBoost (Extreme Gradient Boosting) method, a Classification Technique in ML, and can be illustrated in Ensemble Decision Tree. The reason for selecting the XGBoost method is that it is a high-accuracy method with low computing power. XGboost as part of Supervised Machine Learning, which models the influence of its

independent parameters (**X**) on dependent variables (**Y**) by being integrated into the Ensemble Decision Tree.

The initial stage of research discusses the research phase from data analysis, which starts from the initial analysis (pretest) of data collection results and then data analysis using Structural Equation Modelling (SEM). Furthermore, the researchers found the validity of the contributions of each operational variable using the CFA Method. The next stage is to make predictions using the XGBoost algorithm. Researchers used R software to classify (machine learning) probability with the XGBoost method or algorithm. The output of the model using the XGboost algorithm is the probability value. The XGBoost method is a "Classification Technique" in ML that can be integrated into the ensemble's decision tree. This ensemble method is an algorithm to find the best predictive solution. Ensemble's decision tree combines the predictive values of many of the most straightforward decision trees into one predictive value. With the operational variables selected, the ensembles decision tree using the XGBooT algorithm produces several decision trees for the ERS decision tree model developed in this study. The decision tree model was developed with the XGBoost algorithm requires a maximum of 3 branches.

Considering that this research uses ML, each measurement of the variability is proxied in metric and non-metric. In this study, the dependent variable is dummy, namely committing fraud and not. Moreover, this standard distribution assumption cannot be fulfilled because the independent variable is a mixed variable between continuous variables or metrics and categorical or non-metric. The available Non-Metric or qualitative data are in the form of attributes, characteristics, or categorical traits. Meanwhile, metric or quantitative data is in numbers, where measurements can be made of differences in numbers or degrees. For the measurement scale, the variables are nominal, category, and there is no order. While the nominal scale used is a measurement scale that distinguishes gender, level of education, and others. The following data is in the form of ratios. The measurement of these ratios can be compared, and mathematical operations such as addition, subtraction, multiplication, and division can be performed.

This research begins by investigating and gaining a deeper understanding of fraud in microloans. This study will then predict the behavioral intentions and actual behavior of the current Relationship Manager (RM) to commit fraud in the future. Data is taken from internal data concerning RMs directly involved in providing micro-credit services at bank "X." They have knowledge of service procedures, loan products, and information for potential microloan borrowers. The following hypotheses can be formulated based on the problem formulation and theoretical context mentioned above. This research is based on the fraud triangle theory that the causes of fraud are in the form of three main factors, known as the fraud triangle, the form of Opportunities, Rationalization, and Pressure.

Therefore, this research will address the hypotheses: first, 'Relationship Manager' commits fraud because there is an opportunity. This opportunity arises when the 'Relationship manager' faces ease in the initiation process because there is no supervision from his superiors, and there is a time lag between the initiative and the verdict. These two components can be lowered back into several more complex components (Red flags). The ease of initiative is represented by the number of realization in a certain period. In contrast, the time lag can be represented by the realization in a work unit within a specific time. Second, 'The relationship between the components of rationalization affects the occurrence of fraud.' This rationalization factor can be divided into two main components, namely demographic data and employment data. These demographic red flags are in gender, age, etc., while personal data from employees can be in the form of career data, position, etc. It will be examined whether the Red flags affect future fraud incidents. Third, 'The relationship between Pressure Components affects the occurrence of fraud'. In the pressure component, it will be seen whether the behavior of the financial transaction "Relationship Manager"

has an effect on the possibility of fraud, and whether the performance target is given, and its achievement can also affect the fraud incident. The red flags of the two components include financial condition, lifestyle (transaction profile), and customer management and loan quality targets for the assisted customers.

Lastly, 'Predictive analytics models can predict someone committing fraud in the next 12 months'. The next step is preventive measures; it is necessary to act from the start to predict how likely it is that a "Relationship Manager (RM)" will commit fraud by creating a predictive analytics model. Based on the results of these predictions, personal supervision and guidance of RM can be carried out from the start. The reason for choosing 12 months was that this study was adjusted to the monitoring and evaluation period for achieving the RM performance target, namely 12 months. To run the XGBoost model, the researcher identified the construction components in the variables that affect Opportunity, Rationalization & Pressure. In the econometric equation, the Opportunity, Rationalization, and Pressure hypotheses can encourage fraud. The equation to be written into an ensembles model.

$$\gamma = \underset{c_j \in C.}{\operatorname{argmax}} \sum_{h_i \in H} P(\text{Opportunity, Rationalization, Pressure} | H_i) P(H_i | T) \quad (1)$$

Where " $\gamma$ " is the Fraud Model that will be used to predict, P is the probability of fraud where 'C' is the possibility of fraud in the  $\gamma$ Opportunity, Rationalization & Pressure class and 'H' is the hypothetical for fraud and 'T' is the 'training' of data.

Last part of this research is discussion, validation, and dissemination of results: suggestions and conclusions. In this stage, the validation and dissemination of the results of the predictive analytics model based on variable input / red flags are discussed. Researchers conduct discussion and validation with relevant experts within the company and discuss applying the results of this predictive analytics model in the framework of implementing GRC. The conclusion is determined after the discussion and FGD process has been carried out. This conclusion is the final statement of the dissertation, along with suggestions for further research related to alternative input or additional variables and alternative dissertation processes on the object of research for better research results in the future. The output produced at this stage is the concept of implementing GRC using the results of the predictive analytics model.

## REFERENCES

- Albrecht, S.L. (2012). The influence of job, team and organizational level resources on employee well-being, engagement, commitment and extra-role performance: Test of a model. *International Journal of Manpower*, 33(7), 840–853.
- Andersen, T.B., Bentzen, J., Dalgaard, C.J., & Selaya, P. (2011). Does the internet reduce corruption? Evidence from U.S. states and across countries. *World Bank Economic Review*, 25(3), 387–417.
- Azim, M.I., & Kluvers, R. (2017). Resisting corruption in Grameen bank. *Journal of Business Ethics*, 156(3).
- Baader, G., & Krcmar, H. (2018). Reducing false positives in fraud detection: Combining the red flag approach with process mining. *International Journal of Accounting Information Systems*, 31, 1–16.
- Badinger, H., & Nindl, E. (2014). Globalization and corruption revisited. *The World Economy*, 37(10), 1424–1440.
- Bai, J., Jayachandran, S., Malesky, E.J., & Olken, B. (2013). *Does economic growth reduce corruption? Theory and Evidence*. NBER Working Paper Series.
- Bartlett, J.E., Kotrlik, J.W., & Higgins, C.C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning and Performance Journal*, 19(1), 43–50.
- Baz, R., Samsudin, R.S., Che-Ahmad, A.B., & Popoola, O.M.J. (2016). Capability component of fraud and fraud prevention in the Saudi Arabian banking sector. *International Journal of Economics and Financial Issues*, 6(4), 68–71.

- Bhattacharyya, S., & Hodler, R. (2015). Media freedom and democracy in the fight against corruption. *European Journal of Political Economy*, 39, 13–24.
- Chatterjee, I., & Ray, R. (2012). Does the evidence on corruption depend on how it is measured? Results from a cross-country study on microdata sets. *Applied Economics*, 44(25), 3215–3227.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *KDD 16*, 13-17, 785–794.
- Clarke, R.V. (2012). Opportunity makes the thief. Really? And so what? *Crime Science*, 1(1), 1–9.
- Cohn, A., Maréchal, M.A., & Noll, T. (2010). Bad boys: How criminal identity salience effects rule violation. *Review of Economic Studies*, 82(4), 1289–1308.
- Cressey, D.R. (1954). The differential association theory and compulsive crimes. *The Journal of Criminal Law, Criminology, and Police Science*, 45(1), 29–40.
- Deloitte. (2014). *Fraud risk management – providing insight into fraud prevention, detection, and response*.
- Glaeser, E.L., & Saks, R.E. (2006). Corruption in America. *Journal of Public Economics*, 90(6–7), 1053–1072.
- Goel, R.K., & Nelson, M.A. (2010). Causes of corruption: History, geography, and government. *Journal of Policy Modeling*, 32(4), 433–447.
- Hartungi, R. (2007). Understanding the success factors of micro-finance institutions in a developing country. *International Journal of Social Economics*, 34(6), 388–401.
- Hollow, M. (2014). Money, morals, and motives: An exploratory study into why bank managers and employees commit fraud at work. *Journal of Financial Crime*, 21(2), 174–190.
- Huang, S.Y., Lin, C.C., Chiu, A.A., & Yen, D.C. (2016). Fraud detection using fraud triangle risk factors. *Information Systems Frontiers*, 19(6), 1343–1356.
- Irianto, G., Novianti, N., Rosalina, K., & Firmanto, Y. (2012). Integrity, Unethical Behavior, and Tendency of Fraud. *EKUITAS (Jurnal Ekonomi Dan Keuangan)*, 16(2), 144.
- Lyman, M.D., & Potter, G.W. (2007). Theories of Organized Criminal Behavior. *Organized Crime*, 59–83.
- Mangala, D., & Kumari, P. (2015). Corporate fraud prevention and detection: Revisiting the literature. *Journal of Commerce and Accounting Research*, 4(1).
- Marks, J.T. (2014). Playing offense in a high-risk environment. *Crowe Horwath*, 94(8), 14.
- Meliana, M., & Hartono, T.R. (2019). Fraud Perbankan Indonesia: Studi Eksplorasi. *Prosiding Seminar Nasional Pakar Ke*, 2, 1–7.
- Nawawi, A., & Salin, A.S.A.P. (2018). Employee fraud and misconduct: empirical evidence from a telecommunication company. *Information and Computer Security*, 26(1), 129–144.
- Othman, R., Aris, N.A., Mardiyah, A., Zainan, N., & Amin, N.M. (2015). Fraud detection and prevention methods in the Malaysian public sector: Accountants' and internal auditors' Perceptions. *Procedia Economics and Finance*, 28(April), 59–67.
- Padgett, E.M. (2019). *Behaviors Influencing Accounting Students' Intentions to Whistleblow*.
- Paldam, M., & Gundlach, E. (2011). Two views on institutions and development: The grand transition vs. the primacy of institutions. *SSRN Electronic Journal*.
- PwC. (2020). Fighting fraud: A never-ending battle PwC's Global Economic Crime and Fraud Survey 2020. *PWC Fraud Survey*, 14.
- Suh, J. B., Nicolaides, R., & Trafford, R. (2019). The effects of reducing opportunity and fraud risk factors on the occurrence of occupational fraud in financial institutions. *International Journal of Law, Crime and Justice*, 56, 79–88.
- Treisman, D. (2000). The causes of corruption: A cross-national study. *Journal of Public Economics*, 76, 399–457.
- Van Rijckeghem, C., & Weder, B. (1997). Corruption and the rate of temptation: Do low wages in the civil service cause corruption? *IMF Working Paper*, 97(73), 1.
- Wolfe, D.T., & Hermanson, D.R. (2004). The fraud diamond: Considering the four elements of fraud. *The CPA Journal*, 74(12), 38–42.
- Zamzami, F., Nusa, N.D., & Timur, R.P. (2016). The effectiveness of fraud prevention and detection methods at Universities in Indonesia. *Journal of Research on Leadership Education*, 6(83), 66–69.