

CHAPTER I

INTRODUCTION

1.1 Background

Accounting misstatement is difficult to detect. Misstatement in the financial statements may arise from either fraud or error as stated by International Federations of Accountants (IFAC) [1]. Misleading financial statements will cost a lot, especially for the investors, regulators, and moreover society. It is the auditor's responsibility to provide a reasonable assurance that the financial statements are free from material misstatement.

Enron scandal is one of the most highlighted financial failures in history and literature [2]. It began when Enron shocked the public by booking \$638 million loss. This case dragged its auditor, Arthur Andersen, which failed to detect this, and shredding documents related to Enron audits.

Recent audit controversy involving big 4 accounting firms in Indonesia is when Deloitte Indonesia violates accounting standards principles while auditing PT Sunprima Nusantara Pembiayaan (SNP) caused total losses of \$124 million to the investors in 2018 [3]. Another example from non-big 4 accounting firm is Garuda Indonesia violation where it booked a net profit of \$870,000, when in fact loss of \$175 million is occurred [4], which led to questioning of its auditor, BDO Indonesia.

In environment where the volume and complexity of the data arise, data analytics provide auditors with opportunity to gain a more effective and appropriate understanding of the entity, result in quality of audit risk assessment

[5]. Although data analytics techniques is expected to demonstrate a superior method or at least as supplementary procedure to enrich the current traditional approaches by lowering audit risk, it is seldom or not favorable to use technique such as machine learning on performing audit procedures. This topic is important because while it is possible to be potential contributions to the audit, it is relatively unknown whether usage of data analytics is indeed transformational for the audit [6]. This research conducted experiments on how implementation of neural network may aid detecting fraudulent financial statement, as it may be useful for the earlier audit stage in the planning and risk assessment.

Various research has been conducted in searching of fraudulent financial statement detection, including utilization of supervised learning and unsupervised learning. Supervised learning is used including various models such as Neural Networks [7] - [8], Genetic Algorithm [9], Decision Tree [8] , [10] - [11], Bayesian Network [8], [10], Support Vector Machines [8], [12] - [13], and Logistic Regression [14]. Unsupervised learning implementation used algorithm such as Self-Organizing Map [15] - [16] and K-Means Clustering [15]. This verifies that most of the research in this field are based on supervised learning.

The objective of this research is to develop a supervised learning model that is able to predict fraudulent financial statements from a real public dataset. Financial statements dataset that is publicly available represents severe imbalanced dataset. This is due to the fact that fraudulent financial statement ratio is too small compared to non-fraudulent financial statements. Hence, it is important for a model to able to handle the imbalance problem. Therefore, this study also aims to create a model that is capable of addressing the issue of imbalanced data.

1.2 Identification of Problem

Identified problems on this research are:

- a. The rapid evolution and increasing diversity and complexity of fraud over the years have posed challenges for traditional methods. Due to the vast scale and infrequent occurrence of fraud, it is very difficult to detect similarity of the fraud patterns. Therefore, a solution that can learn from the complexity of the fraud patterns is crucial. This leads to necessities of utilizing sophisticated techniques like Machine Learning or Neural Networks which show promising results in identifying complex patterns effectively.
- b. In many instances, non-fraudulent financial statements can extremely outnumber the fraudulent financial statement which creates an imbalance dataset scenario. This scenario poses a challenge as it can lead the model to “under-learn” the minority class (in this case, the fraudulent financial statements) which can impair the performance of the model to correctly identify the fraud. This can result in a higher rate of false positives or false negatives. Therefore, it is important to develop a model that is able to counter the imbalance issue.
- c. While existing studies have demonstrated the capabilities of various models to detect fraudulent financial statements, there are still areas that can be explored. The most critical aspect of the improvement point is accuracy. Hence, developing a model with better accuracy is crucial so it can be relied upon especially for practical usage.

1.3 Research Scope

The range of the observation in the research is limited to:

- a. This research will be dedicated to the fraud detection area. Specifically, the research is limited to the study of detection of fraudulent financial statements. Hence, the aim of the research is to experiment and evaluate the performance of applied proposed model in identifying fraudulent financial statements.
- b. The research will use financial statements from the dataset of the previous study by Bao et al. [13] representing public listed U.S. firms within the range period of 1990-2014. Thus, the results of this study can be benchmarked with the results of the previous study.
- c. This study will specifically explore the applications of Neural Networks, specifically, Artificial Neural Network (ANN). Following the identified problems of fraud cases which are becoming more diverse and complex, for this reason, this study aims to leverage ANN to address the problems in complex and non-linear characteristics of fraudulent financial statements.

1.4 Problem Formulation

This research attempted to investigate and answer:

- a. Among developed Artificial Neural Networks, featuring different fraud detection predictors, which one demonstrates superior performance in detecting fraudulent financial statements?

- b. What is the effectiveness of the proposed model in terms of accuracy of detecting fraudulent financial statements in a severe imbalance dataset?

1.5 Research Objectives

The objective of this study is to develop a model that is able to demonstrate the capabilities of detecting fraudulent financial statements within publicly available dataset. The focus will be on the best model that produces the best results, which indicates the effectiveness of the model to handle real-world data. In addition, this research aims to benchmark the results produced with previous results in terms of accuracy.

1.6 Thesis Outline

This thesis will be divided into 5 chapters, which each chapter will discuss on specific and different purpose. The systematics writing of this thesis includes:

Chapter I Introduction. This chapter briefly discusses the needs of fraudulent financial statements detection, challenge and problem in creating one, scope and limitations in this study, and the research objectives.

Chapter II Literature Review. This chapter presents comprehensive literature review on fraudulent financial statements detection and underlying theories of proposed fraudulent financial statements detection model which covered fraud predictors, ANN, and SMOTE.

Chapter III Research Methodology. This chapter presents the research workflow, the dataset used along with the relevant preprocessing steps that consists of serial fraud treatment, SMOTE application to combat imbalance issue, and the

detail of the experiments conducted which consists of three experiments on different fraud predictors aiming to address the defined problems.

Chapter IV Results and Discussion. This chapter elaborates on the results of each experiment conducted in this study and benchmarked the results with previous fraudulent financial statements study. This chapter will attach the related published paper.

Chapter V Conclusion. This chapter presents the conclusion based on the results obtained and provides any constructive suggestion that may be used for the next research and its utility for practitioners.

At the end of this writing include references, appendices, and author biography.