automate and improve the accuracy of fraud detection by distinguishing fraudulent activities from legitimate ones [7]. Machine learning algorithms learn to tell fraudulent operations from legitimate ones without raising the suspicions of those executing the transactions [8],[9].

There are many research or studies that have been conducted related to fraud detection using machine learning. Some of those are as follows:

a. Bank Fraud Detection Using Support Vector Machine[10] : This paper examines the different manifestations of fraudulent activities within banking institutions, employing data mining techniques to facilitate the timely identification of such activities by leveraging existing data stored within the bank's systems. Supervised learning approaches, specifically Support Vector Machines with Spark (SVM-S), are employed to construct models that capture both regular and deviant client behaviour. These models are subsequently utilised to assess the legitimacy of novel transactions. The efficacy of these strategies in combating banking fraud in big data is supported by the findings derived from credit card transaction databases. The experimental findings of this study indicate that Support Vector Machines with Sequential Minimal Optimisation (SVM-S) have superior predictive capabilities compared to Back Propagation Networks (BPN). In addition to the mean prediction, the highest level of accuracy is achieved when the training data ratio exceeds 0.8.
b. Fraud Detection suspicion on credit cards with Machine Learning[11] : This study tried to implement the Random Forest method to find the best solution

to predict the credit card fraud and obtained a result of 0.85 which indicates that the developed model has good accuracy.

c. Comparison of Seven ATM Fraud Identification Algorithms at PT. Bank Central Asia Tbk[12] : In this study, the investigator gathered a total of five datasets and performed preprocessing procedures on the collected data. This preprocessing was conducted to render the dataset suitable for subsequent modelling and algorithmic testing, with the aim of addressing the encountered research questions. The study encompassed a total of seven algorithmic comparisons, specifically focusing on Decision Trees, Gradient Boosted Trees, Logistic Regression, Naive Bayes (Kernels), Random Forest, and Random Tree. Based on the conducted modelling and testing, the findings indicate that the Gradient Boosted Trees algorithm exhibits superior performance. It achieves an accuracy rate of 99.85% and an AUC value of 1. The exceptional performance of this algorithm can be attributed to its inherent compatibility with the attributes tested, which aligns with the characteristics and capabilities of the Gradient Boosted Trees algorithm. Consequently, the algorithm demonstrates robustness and consistently produces reliable outcomes. The Gradient Boosted Trees algorithm has been employed as a solution for the detection of ATM machine fraud in PT. Bank Central Asia Tbk.

d. Credit Card Fraud Detection Using Machine Learning Final Research Paper [13] : This study employed the Local Outlier Factor, K-Nearest Neighbours (KNN), and Random Forest algorithms for the purpose of fraud detection. The evaluation of the model's performance is conducted by considering metrics

such as accuracy, sensitivity, precision, and recall. The outcome indicates that Random Forest yields more precise predictions and necessitates time for both the training and testing phases. The random forest algorithm achieves a high level of accuracy, specifically 99.980%.

e. Random Forest For Credit Card Fraud Detection[14] : This study utilises two variations of the Random Forest algorithm to teach the behavioural characteristics of both normal and aberrant transactions. The authors conduct a comparative analysis of two distinct Random Forest models that differ in their base classifiers, evaluating their respective performance in the context of credit card fraud detection. The data utilised in this experiment originates from a Chinese e-commerce corporation.

Nevertheless, the implementation of machine learning in the context of fraud detection requires thorough examination in order to determine the most optimal approaches. The primary objective of this work is to address the existing disparity and augment the capabilities of fraud detection through an investigation into the efficacy of three widely used machine learning models, namely Support Vector Machine, Random Forest, and Gradient Boosting, in the context of identifying instances of banking fraud.

1.2 Problem Identification

Currently, fraud in bank transactions at Bank is difficult to detect because it is highly dependent on the expertise of auditors. To recognize the indicator of fraud found in financial statements or existing transaction data, auditors must be equipped with sufficient knowledge to analyze data samples which often developed through

lengthy period of training and experience. On top of that, the transaction data size in this case study is considerably massive, and it is possible that the transaction data sample taken does not contain fraud. At times, fraud is detected when a customer or employee reports it. To overcome limitations in detecting frauds, banks need new efforts or methods to detect them quickly and accurately.

1.3 Limitations of The Problem

1.3.1 Methods

The study will use three primary machine learning algorithms for model development: Support Vector Machines (SVM), Random Forest, and Gradient Boosting. Each model will be trained and fine-tuned using the bank's historical transaction data, then tested using a withheld portion of the data to evaluate its performance. The study will also explore the feature importance provided by these algorithms to understand which factors contribute most significantly to fraud detection.

Model performance will be evaluated using various metrics such as the confusion matrix, CA, precision, recall, F1-Score, ROC curve, and AUC. These metrics will provide a comprehensive understanding of the performance of each model, allowing for a meaningful comparison between them.

1.3.2 Data

The data used for this study will come from real-world ATM transactions obtained from a Regional Development Bank in Indonesia. It comprises 877,683 transactions, including both fraudulent and legitimate transactions. The dataset has

been pre-processed and includes 44 attribute values such as transaction date, transaction time, ATM terminal id, transaction code, and transaction amount.. A new feature called "TARGET_DATA" will be introduced to label transactions as legitimate (0) or fraudulent (1). This will allow the study to explore the patterns and characteristics of fraudulent transactions.

1.3.3 Scope

This research will be specifically focus on skimming fraud in ATM transaction at Regional Development Bank in Indonesia. It will examine the effectiveness of machine learning algorithms for fraud detection within this specific context, aiming to address the unique needs and challenges of the bank.

While the research is intended to be applicable to the regional bank, the insights and methodologies may also be of value to other banks or financial institutions grappling with similar challenges. However, given the specificity of the dataset and the context, direct application to other contexts may require additional modifications and considerations.

1.4 Problem Formulation

Based on the previously mentioned research, three specific machine learning techniques have been recognised as viable instruments for the detection of fraudulent activities. These techniques are Support Vector Machine, Random Forest, and Gradient Boosted Trees. Acknowledging the imperative to augment the efficacy and precision of fraud detection within the banking industry, alongside the notable relevance of employing machine learning methodologies in this domain, we

have devised vital research inquiries. The purpose of these inquiries is to analyse the efficacy and relative performance of the machine learning models in issue. Consequently, we formulated the following study queries:

1. To what extent are machine learning methods, specifically Support Vector Machine, Random Forest, and Gradient Boosting, effective in detecting fraudulent ATM transactions within the banking data?
2. What is the comparative evaluation of the three investigated machine learning methods (Support Vector Machine, Random Forest, and Gradient Boosting) in the context of detecting skimming fraud in ATM transactions, specifically within the operational framework of a Regional Development Bank in Indonesia, based on metrics such as the confusion matrix, accuracy, sensitivity, precision, recall, ROC Curve, and AUC as presented in the case study of this thesis?

1.5 Research Purposes

The objective of this study is to conduct a thorough comparative examination of several machine learning techniques for the identification of skimming fraud in Automated Teller Machine (ATM) transactions. The objective of this study is to evaluate the efficacy of three distinct machine learning models, namely Support Vector Machine, Random Forest, and Gradient Boosting. The primary aim is to determine which of these models produces the most precise and efficient outcomes. The objective of this study is to assess the performance of several models in detecting fraudulent ATM transactions by analysing their accuracy, precision, sensitivity, and recall metrics. The primary objective is to

utilise the capabilities of machine learning in order to augment the existing fraud detection systems, thereby offering clients with financial services that are safer and more secure.

1.6 Thesis Outline

The writing in this thesis is divided into at least five chapters, where each chapter has a discussion of different goals and contents. The systematics are as follows:

Chapter I Introduction. This chapter discusses a brief description of the background of the problem why this research was carried out to reach the research objectives.

Chapter II Literature Review. This chapter discusses the theories that will be used, or the research that has been carried out related to the formulation of the problems discussed in Chapter 1. This section is the key part to determine the method that will be used in the next section.

Chapter III Research Methodology. This chapter contains the research design or test design .

Chapter IV Result and Analysis. Describe the results of the research that has been carried out and make an argument for what is produced.

Chapter V Conclusion. This chapter describes the conclusions based on the results of the research obtained, as well as constructive suggestions that need to be developed for future research so that the next research will be better.

At the end of this paper, a bibliography, appendices and curriculum vitae of the researcher are attached.