

CHAPTER 1

INTRODUCTION

1.1. Background

Fama [1] described that it is impossible to predict stock prices because of its own behavior even though it has efficient information markets. There are 2 ways to approach the market which are Fundamental Analysis and Technical Analysis and both of them shows its own limit to help people to beat the market. According to Investopedia, Technical Analysis has its own flaws such as chart pattern that does not guarantee 100% works all the time and other variable such as news can break the past record and Fundamental Analysis has some disadvantages such as financial engineering to fake financial engineering and dishonest management like what happened in Enron where many investor lost their money. But people are still confused to choose which strategies that are best to use for. Therefore i want to choose one technique and make simplified version of technical analysis. Stock Market in Indonesia has been perceived as place to gamble and bad place for investment prior to many incident such as mutual fund who goes bankrupt, stock manipulation in recent incident on ASABRI and Jiwasraya. Heru Hidayat and Benny Tjokrosaputro found guilty for corruption in that case [2]. The other issue are people who does not have time to study or learn how to invest can not really trust stock analyst. Investor who have bought Bukalapak stock (BUKA) because of the forecast feel disappointed. Because after reaching 1325 on the second day of IPO, the price fell down sharply. Even Nomura Securities who forecasted the price from Rp.1620 to Rp.560 [3] With this concern, the writer believes that stock prices that run with machine learning and arithmetic will gain investor trust again and perform better because it does not have any attachment with emotion.

Writer comes up with solution that revolves around Artificial Neural Network (ANN) and Random Forest and Arima (Autoregressive Integrated Moving Average) model to predict stock prices by evaluating data from Jakarta Stock Exchange taken from Yahoo Finance. There are data such as Open, High, Close, Low, Volume, Adjusted Close, Year, Month, Day with 10 years period that been investigate in this paper. Open represents the opening price of the day in the chart, High is the highest price of the day in the chart, Close means the closing price of the day in the chart, Low is the lowest price of the day, Volume represents the number of shares traded in a stock in that

day, Adjusted close represents the closing price after adjustments for all applicable splits and dividend distributions, and Year, Month and Day represent the date of the stock in the chart. Writer created model that will be evaluated with Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE).

1.2. Problem Formulation

Based on the issue and the circumstances, the writer aims to solve those issues from a science perspective regarding machine learning. There are some questions that inspire the writer to investigate this topic. These are some of the questions :

1. Which machine learning model performs better for coal stocks?
2. Which machine learning has the lowest error percentage?

1.3. Scope

Writer uses 10 years period of data coal stocks from Yahoo Finance that contains Open, High, Close, Low, Volume, Adjusted Close, Year, Month, Day. It starts from First of January 2010 and ends in 31st of December 2020. The writer compares the result from Random Forest, Arima (Autoregressive Integrated Moving Average), and Artificial Neural Network (ANN) in stock and which one performs the best.

1.4. Objective

The purpose of this project is to evaluate three machine learning models in predicting specific coal stocks in Indonesia Stock Exchange. Another objective is to find machine learning model with the lowest error percentage by comparing the three models (Random Forest, Arima (Autoregressive Integrated Moving Average), and Artificial Neural Network (ANN)) and those models will be evaluated with some metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE).